



US009096871B2

(12) **United States Patent**  
Stege et al.

(10) **Patent No.:** US 9,096,871 B2  
(45) **Date of Patent:** Aug. 4, 2015

(54) **VARIANT CBH I POLYPEPTIDES WITH REDUCED PRODUCT INHIBITION**

(75) Inventors: **Justin T. Stege**, San Diego, CA (US); **Alexander Varvak**, Netanya (IL); **John Poland**, San Diego, CA (US); **Chris S. Lyon**, San Diego, CA (US); **Shaun Healey**, San Diego, CA (US); **Peter Luginbuhl**, San Diego, CA (US)

(73) Assignee: **BP Corporation North America Inc.**, Houston, TX (US)

(\*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 158 days.

(21) Appl. No.: **13/824,317**

(22) PCT Filed: **Oct. 6, 2011**

(86) PCT No.: **PCT/US2011/055181**

§ 371 (c)(1),  
(2), (4) Date: **Dec. 18, 2013**

(87) PCT Pub. No.: **WO2012/048171**

PCT Pub. Date: **Apr. 12, 2012**

(65) **Prior Publication Data**

US 2014/0147894 A1 May 29, 2014

**Related U.S. Application Data**

(60) Provisional application No. 61/390,392, filed on Oct. 6, 2010.

(51) **Int. Cl.**

**C12N 9/42** (2006.01)  
**C12P 7/14** (2006.01)

**C12P 19/02** (2006.01)  
**C12P 19/14** (2006.01)

(52) **U.S. Cl.**  
CPC ..... **C12P 7/14** (2013.01); **C12N 9/2437** (2013.01); **C12P 19/02** (2013.01); **C12P 19/14** (2013.01); **C12Y 302/01091** (2013.01); **Y02E 50/16** (2013.01); **Y02E 50/17** (2013.01)

(58) **Field of Classification Search**

None  
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

8,101,393 B2 \* 1/2012 Gray et al. .... 435/209  
2009/0162916 A1 6/2009 Adney

FOREIGN PATENT DOCUMENTS

EP 2357227 8/2011  
WO WO 2004/078919 9/2004

\* cited by examiner

*Primary Examiner* — Nashaat Nashed

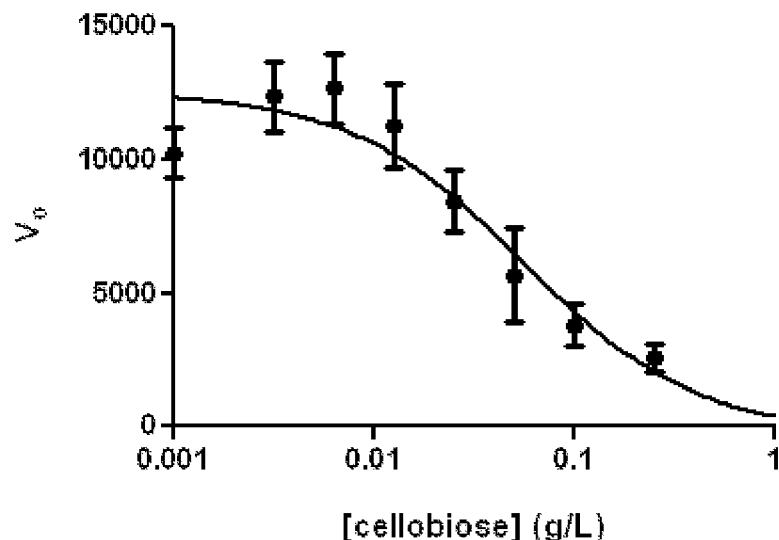
(74) *Attorney, Agent, or Firm* — DLA Piper LLP (US)

(57) **ABSTRACT**

The present disclosure relates to variant CBH I polypeptides that have reduced product inhibition, and compositions, e.g., cellulase compositions, comprising variant CBH I polypeptides. The variant CBH I polypeptides and related compositions can be used in variety of agricultural and industrial applications. The present disclosure further relates to nucleic acids encoding variant CBH I polypeptides and host cells that recombinantly express the variant CBH I polypeptides.

**57 Claims, 3 Drawing Sheets**

A.



B.

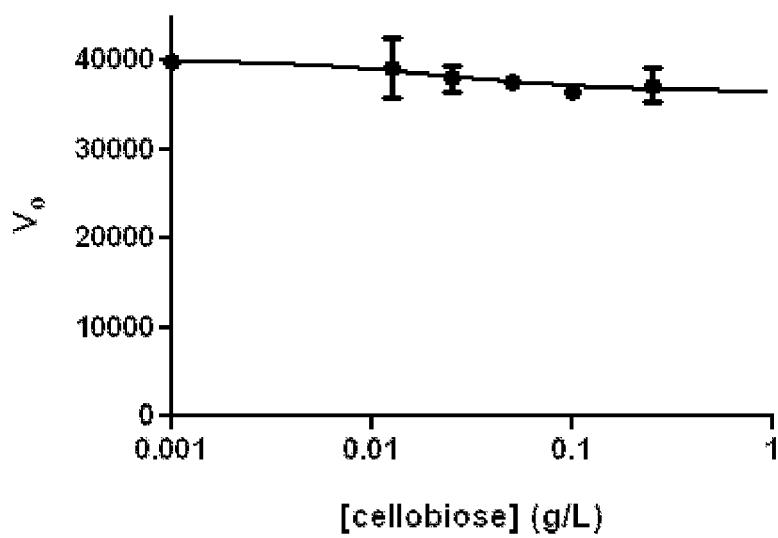
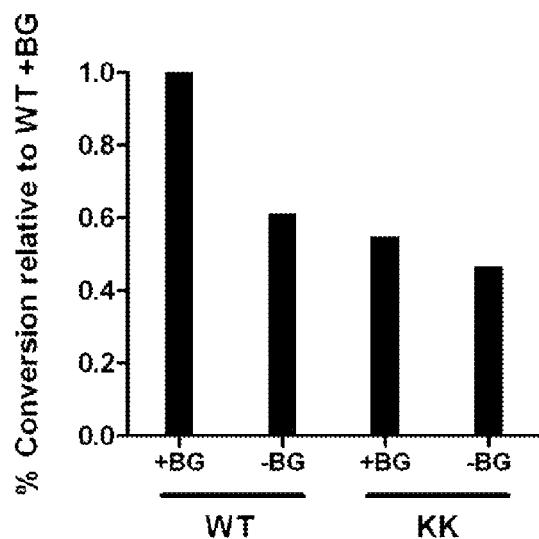


FIGURE 1A-1B

A.



B.

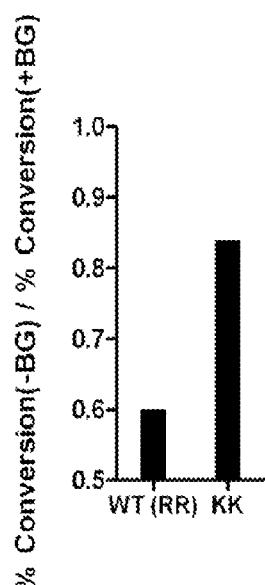


FIGURE 2A-2B

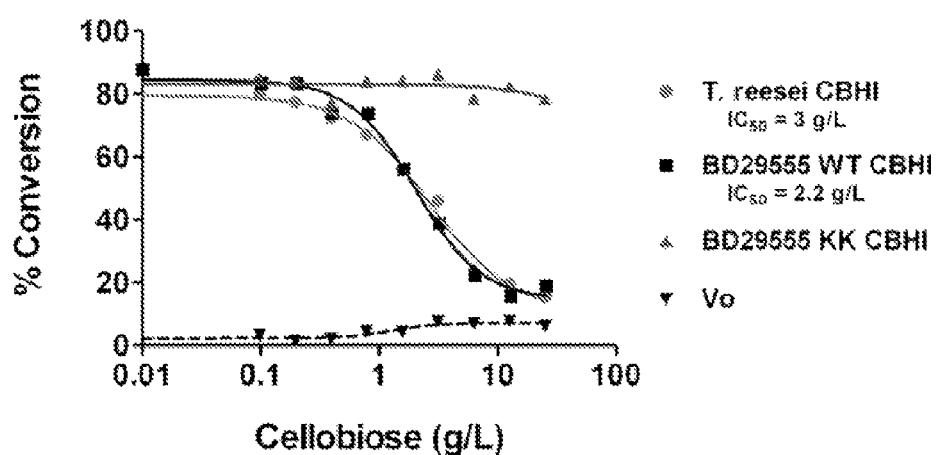


FIGURE 3

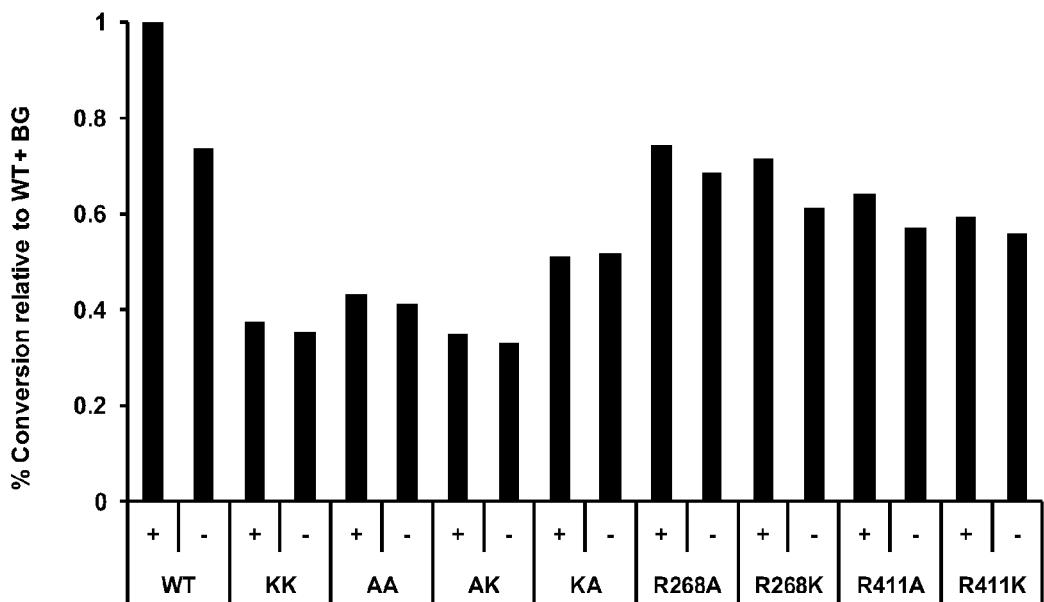


FIGURE 4

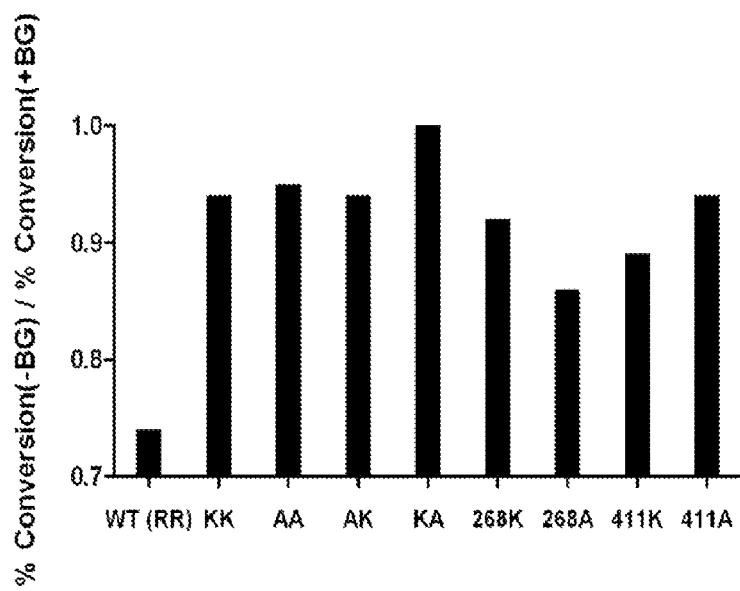


FIGURE 5

**1****VARIANT CBH I POLYPEPTIDES WITH REDUCED PRODUCT INHIBITION****CROSS-REFERENCE TO RELATED APPLICATIONS**

The present application claims the benefit of national stage application filed in compliance with 35 U.S.C. §371 of International Application No. PCT/US2011/055181, filed Oct. 6, 2011, which claims benefit under 35 U.S.C. §119(3) of U.S. Provisional Application No. 61/390,392, filed Oct. 6, 2010, and are herein incorporated in their entireties for all purposes.

**BACKGROUND**

Cellulose is an unbranched polymer of glucose linked by  $\beta(1 \rightarrow 4)$ -glycosidic bonds. Cellulose chains can interact with each other via hydrogen bonding to form a crystalline solid of high mechanical strength and chemical stability. The cellulose chains are depolymerized into glucose and short oligosaccharides before organisms, such as the fermenting microbes used in ethanol production, can use them as metabolic fuel. Cellulase enzymes catalyze the hydrolysis of the cellulose (hydrolysis of  $\beta$ -1,4-D-glucan linkages) in the biomass into products such as glucose, cellobiose, and other cellooligosaccharides. Cellulase is a generic term denoting a multienzyme mixture comprising exo-acting cellobiohydrolases (CBHs), endoglucanases (EGs) and  $\beta$ -glucosidases (BGs) that can be produced by a number of plants and microorganisms. Enzymes in the cellulase of *Trichoderma reesei* include CBH I (more generally, Cel7A), CBH2 (Cel6A), EG1 (Cel7B), EG2 (Cel5), EG3 (Cel12), EG4 (Cel61A), EG5 (Cel45A), EG6 (Cel74A), Cip1, Cip2,  $\beta$ -glucosidases (including, e.g., Cel3A), acetyl xylan esterase,  $\beta$ -mannanase, and swollenin.

Cellulase enzymes work synergistically to hydrolyze cellulose to glucose. CBH I and CBH II act on opposing ends of cellulose chains (Barr et al., 1996, Biochemistry 35:586-92), while the endoglucanases act at internal locations in the cellulose. The primary product of these enzymes is cellobiose, which is further hydrolyzed to glucose by one or more  $\beta$ -glucosidases.

The cellobiohydrolases are subject to inhibition by their direct product, cellobiose, which results in a slowing down of saccharification reactions as product accumulates. There is a need for new and improved cellobiohydrolases with improved productivity that maintain their reaction rates during the course of a saccharification reaction, for use in the conversion of cellulose into fermentable sugars and for related fields of cellulosic material processing such as pulp and paper, textiles and animal feeds.

**SUMMARY**

The present disclosure relates to variant CBH 1 polypeptides. Most naturally occurring CBH 1 polypeptides have arginines at positions corresponding to R268 and R411 of *T. reesei* CBH I (SEQ ID NO:2). The variant CBH 1 polypeptides of the present disclosure include a substitution at either or both positions resulting in a reduction or decrease in product (e.g., cellobiose) inhibition. Such variants are sometimes referred to herein as "product tolerant."

The variant CBH 1 polypeptides of the disclosure minimally contain at least a CBH I catalytic domain, comprising (a) a substitution at the amino acid position corresponding to R268 of *T. reesei* CBH I ("R268 substitution"); (b) a substitution at the amino acid position corresponding to R411 of *T.*

**2**

*reesei* CBH I ("R411 substitution"); or (c) both an R268 substitution and an R411 substitution. The amino acid positions of exemplary CBH I polypeptides into which R268 and/or R411 substitutions can be introduced are shown in Table 1, and the amino acid positions corresponding to 8268 and/or R411 in these exemplary CBH I polypeptides are shown in Table 2.

R268 and/or R411 substituents can include lysines and/or alanines. Accordingly, the present disclosure provides a variant CBH I polypeptide comprising a CBH I catalytic domain with one of the following amino acid substitutions or pairs of R268 and/or R411 substitutions: (a) R268K and R411K; (b) R268K and R411A; (c) R268A and R411K; (d) R268A and R411A; (e) R268A; (f) R268K; (g) R411A; and (h) R411K. In some embodiments, however, the amino acid sequence of the variant CBH I polypeptide does not comprise or consist of SEQ ID NO:299, SEQ ID NO:300, SEQ ID NO:301, or SEQ ID NO:302.

The variant CBH I polypeptides of the disclosure typically include a CD comprising an amino acid sequence having at least 50% sequence identity to a CD of a reference CBH I exemplified in Table 1. The CD portions of the CBH I polypeptides exemplified in Table 1 are delineated in Table 3. The variant CBH I polypeptides can have a cellulose binding domain ("CBD") sequence in addition to the catalytic domain ("CD") sequence. The CBD can be N- or C-terminal to the CD, and the CBD and CD are optionally connected via a linker sequence.

The variant CBH I polypeptides can be mature polypeptides or they may further comprise a signal sequence.

Additional embodiments of the variant CBH I polypeptides are provided in Section 5.1.

The variant CBH I polypeptides of the disclosure typically exhibit reduced product inhibition by cellobiose. In certain embodiments, the IC<sub>50</sub> of cellobiose towards a variant CBH I polypeptide of the disclosure is at least 1.2-fold, at least 1.5-fold, or at least 2-fold the IC<sub>50</sub> of cellobiose towards a reference CBH I lacking the R268 substitution and/or R411 substitution present in the variant. Additional embodiments of the product inhibition characteristics of the variant CBH I polypeptides are provided in Section 5.1.

The variant CBH I polypeptides of the disclosure typically retain some cellobiohydrolase activity. In certain embodiments, a variant CBH I polypeptide retains at least 50% the CBH I activity of a reference CBH I lacking the R268 substitution and/or R411 substitution present in the variant. Additional embodiments of cellobiohydrolase activity of the variant CBH I polypeptides are provided in Section 5.1.

The present disclosure further provides compositions (including cellulase compositions, e.g., whole cellulase compositions, and fermentation broths) comprising variant CBH I polypeptides. Additional embodiments of compositions comprising variant CBH I polypeptides are provided in Section 5.3. The variant CBH I polypeptides and compositions comprising them can be used, inter alia, in processes for saccharifying biomass. Additional details of saccharification reactions, and additional applications of the variant CBH I polypeptides, are provided in Section 5.4.

The present disclosure further provides nucleic acids (e.g., vectors) comprising nucleotide sequences encoding variant CBH I polypeptides as described herein, and recombinant cells engineered to express the variant CBH I polypeptides. The recombinant cell can be a prokaryotic (e.g., bacterial) or eukaryotic (e.g., yeast or filamentous fungal) cell. Further

provided are methods of producing and optionally recovering the variant CBH I polypeptides. Additional embodiments of the recombinant expression system suitable for expression and production of the variant CBH I polypeptides are provided in Section 5.2.

#### BRIEF DESCRIPTION OF THE FIGURES AND TABLES

FIG. 1A-1B: Cellobiose dose-response curves using a 4-MUL assay for a wild-type CBH I (BD29555; FIG. 1A) and a R268K/R411K variant CBH I (BD29555 with the substitutions R273K/R422K; FIG. 1B).

FIG. 2A-2B: The effect of cellobiose accumulation on the activity of wild-type CBH I and a R268K/R411K variant CBH I, based on percent conversion of glucan after 72 hours in the bagasse assay. FIG. 2A shows relative activity in the presence (+) and absence (-) of  $\beta$ -glucosidase (BG), where relative activity is normalized to wild type activity with BG (WT+=1). FIG. 2B shows tolerance to cellobiose as a function of the ratio of activity in the absence vs. presence of  $\beta$ -glucosidase (activity ratio=Activity-BG/Activity+BG).

FIG. 3: Cellobiose dose-response curves using PASC assay for a R268K/R411K variant CBH I polypeptide as compared to two wild type CBH I polypeptides.

FIG. 4: The effect of cellobiose accumulation on the activity of a wild-type CBH I and a R268K/R411K variant CBH I based on percent conversion of glucan after 72 hours in the bagasse assay in the presence (+) and absence (-) of  $\beta$ -glucosidase (BG). Activity is normalized to wild type activity with BG (WT+=1).

FIG. 5: Characterization of cellobiose product tolerance of variant CBH I polypeptides, based on percent conversion of glucan after 72 hours in the absence and presence of  $\beta$ -glucosidase (BG) in the bagasse assay; tolerance is evaluated as a function of the ratio of activity in the absence vs. presence of  $\beta$ -glucosidase.

TABLE 1: Amino acid sequences of exemplary “reference” CBH I polypeptides that can be modified at positions corresponding to R268 and/or R411 in *T. reesei* CBH I (SEQ ID NO:2). The database accession numbers are indicated in the second column. Unless indicated otherwise, the accession numbers refer to the Genbank database. “#” indicates that the CBH I has no signal peptide; “&” indicate that the sequence is from the PDB database and represents the catalytic domain only without signal sequence; \* indicates a nonpublic database. These amino acid sequences are mostly wild type, with the exception of some sequences from the PDB database which contain mutations to facilitate protein crystallization.

TABLE 2: Amino acid positions in the exemplary reference CBH I polypeptides that correspond to R268 and R411 in *T. reesei* CBH I. Database descriptors are as for Table 1.

TABLE 3: Approximate amino acid positions of CBH I polypeptide domains. Abbreviations used: SS is signal sequence; CD is catalytic domain; and CBD is cellulose binding domain. Database descriptors are as for Table 1.

TABLE 4: Table 4 shows a segment within the catalytic domain of each exemplary reference CBH I polypeptide containing the active site loop (shown in bold, underlined text) and the catalytic residues (glutamates in most CBH I polypeptides) (shown in bold, double underlined text). Database descriptors are as for Table 1.

TABLE 5: MUL and bagasse assay results for variants of BD29555. ND means not determined.  $\pm\%$  Activity (+/-cellobiose)=[(Activity with cellobiose)/(Activity without cellobiose)]\*100.  $\mp\%$  Activity (-/+BG)=[(Activity without BG)/(Activity with BG)]\*100

TABLE 6: MUL and bagasse assay results for variants of *T. reesei* CBH I. ND means not determined.  $\pm\%$  Activity (+/-cellobiose)=[(Activity with cellobiose)/(Activity without cellobiose)]\*100.  $\mp\%$  Activity (-/+BG)=[(Activity without BG)/(Activity with BG)]\*100.

TABLE 7: Informal sequence listing. SEQ ID NO:1-149 correspond to the exemplary reference CBH I polypeptides. SEQ ID NO:299 corresponds to mature *T. reesei* CBH I (amino acids 26-529 of SEQ ID NO:2) with an R268A substitution. SEQ ID NO:300 corresponds to mature *T. reesei* CBH I (amino acids 26-529 of SEQ ID NO:2) with an R411A substitution. SEQ ID NO:301 corresponds to full length BD29555 with both an R268K substitution and an R411K substitution. SEQ ID NO:302 corresponds to mature BD29555 with both an R268K substitution and an R411K substitution. SEQ ID NO:152 corresponds to full length BD29555 with both an R268K substitution and an R411K substitution. SEQ ID NO:153 corresponds to mature BD29555 with both an R268K substitution and an R411K substitution.

#### DETAILED DESCRIPTION

The present disclosure relates to variant CBH I polypeptides. Most naturally occurring CBH I polypeptides have arginines at positions corresponding to R268 and R411 of *T. reesei* CBH I (SEQ ID NO:2). The variant CBH I polypeptides of the present disclosure include a substitution at either or both positions resulting in a reduction of product (e.g., cellobiose) inhibition. The following subsections describe in greater detail the variant CBH I polypeptides and exemplary methods of their production, exemplary cellulase compositions comprising them, and some industrial applications of the polypeptides and cellulase compositions.

##### Variant CBH I Polypeptides

The present disclosure provides variant CBH I polypeptides comprising at least one amino acid substitution that results in reduced product inhibition. “Variant” means a polypeptide which differs in sequence from a reference polypeptide by substitution of one or more amino acids at one or a number of different sites in the amino acid sequence. Exemplary reference CBH I polypeptides are shown in Table 1.

The variant CBH I polypeptides of the disclosure have an amino acid substitution at the amino acid position corresponding to R268 of *T. reesei* CBH I (SEQ ID NO:2) (an “R268 substitution”), (b) a substitution at the amino acid position corresponding to R411 of *T. reesei* CBH I (“R411 substitution”); or (c) both an R268 substitution and an R411 substitution, as compared to a reference CBH I polypeptide. It is noted that the R268 and R411 numbering is made by reference to the full length *T. reesei* CBH I, which includes a signal sequence that is generally absent from the mature enzyme. The corresponding numbering in the mature *T. reesei* CBH I (see, e.g., SEQ ID NO:4) is R251 and R394, respectively.

Accordingly, the present disclosure provides variant CBH I polypeptides in which at least one of the amino acid positions corresponding to R268 and R411 of *T. reesei* CBH I, and optionally both the amino acid positions corresponding to R268 and R411 of *T. reesei* CBH I, is not an arginine.

The amino acid positions in the reference polypeptides of Table 1 that correspond to R268 and R411 in *T. reesei* CBH I are shown in Table 2. Amino acid positions in other CBH I polypeptides that correspond to R268 and R411 can be identified through alignment of their sequences with *T. reesei*

CBH I using a sequence comparison algorithm. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, 1981, *Adv. Appl. Math.* 2:482-89; by the homology alignment algorithm of Needleman & Wunsch, 1970, *J. Mol. Biol.* 48:443-53; by the search for similarity method of Pearson & Lipman, 1988, *Proc. Nat'l Acad. Sci. USA* 85:2444-48, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr, Madison, Wis.), or by visual inspection.

The R268 and/or R411 substitutions are preferably selected from (a) R268K and R411K; (b) R268K and R411A; (c) R268A and R411K; (d) R268A and R411A; (e) R268A; (f) R268K; (g) R411A; and (h) R411K.

CBH I polypeptides belong to the glycosyl hydrolase family 7 ("GH7"). The glycosyl hydrolases of this family include endoglucanases and cellobiohydrolases (exoglucanases). The cellobiohydrolases act processively from the reducing ends of cellulose chains to generate cellobiose. Cellulases of bacterial and fungal origin characteristically have a small cellulose-binding domain ("CBD") connected to either the N or the C terminus of the catalytic domain ("CD") via a linker peptide (see Suomalaki et al., 2000, *Cellulose* 7: 189-209). The CD contains the active site whereas the CBD interacts with cellulose by binding the enzyme to it (van Tilbeurgh et al., 1986, *FEBS Lett.* 204(2): 223-227; Tomme et al., 1988, *Eur. J. Biochem.* 170:575-581). The three-dimensional structure of the catalytic domain of *T. reesei* CBH I has been solved (Divne et al., 1994, *Science* 265:524-528). The CD consists of two  $\beta$ -sheets that pack face-to-face to form a  $\beta$ -sandwich. Most of the remaining amino acids in the CD are loops connecting the  $\beta$ -sheets. Some loops are elongated and bend around the active site, forming cellulose-binding tunnel of (~50 Å). In contrast, endoglucanases have an open substrate binding cleft/groove rather than a tunnel. Typically, the catalytic residues are glutamic acids corresponding to E229 and E234 of *T. reesei* CBH I.

The loops characteristic of the active sites ("the active site loops") of reference CBH I polypeptides, which are absent from GH7 family endoglucanases, as well as catalytic glutamate residues of the reference CBH I polypeptides, are shown in Table 4. The variant CBH I polypeptides of the disclosure preferably retain the catalytic glutamate residues or may include a glutamine instead at the position corresponding to E234, as for SEQ ID NO:4. In some embodiments, the variant CBH I polypeptides contain no substitutions or only conservative substitutions in the active site loops relative to the reference CBH I polypeptides from which the variants are derived.

Many CBH I polypeptides do not have a CBD, and most studies concerning the activity of cellulase domains on different substrates have been carried out with only the catalytic domains of CBH I polypeptides. Because CDs with cellobiohydrolase activity can be generated by limited proteolysis of mature CBH I by papain (see, e.g., Chen et al., 1993, *Biochem. Mol. Biol. Int.* 30(5):901-10), they are often referred to as "core" domains. Accordingly, a variant CBH I can include only the CD "core" of CBH I. Exemplary reference CDs comprise amino acid sequences corresponding to positions 26 to 455 of SEQ ID NO:1, positions 18 to 444 of SEQ ID NO:2, positions 26 to 455 of SEQ ID NO:3, positions 1 to 427 of SEQ ID NO:4, positions 24 to 457 of SEQ ID NO:5, positions 18 to 448 of SEQ ID NO:6, positions 27 to 460 of SEQ ID NO:7, positions 27 to 460 of SEQ ID NO:8, positions 20 to 449 of SEQ ID NO:9, positions 1 to 424 of SEQ ID NO:10, positions 18 to 447 of SEQ ID NO:11, positions 18 to

434 of SEQ ID NO:12, positions 18 to 445 of SEQ ID NO:13, positions 19 to 454 of SEQ ID NO:14, positions 19 to 443 of SEQ ID NO:15, positions 2 to 426 of SEQ ID NO:16, positions 23 to 446 of SEQ ID NO:17, positions 19 to 449 of SEQ ID NO:18, positions 23 to 446 of SEQ ID NO:19, positions 19 to 449 of SEQ ID NO:20, positions 2 to 416 of SEQ ID NO:21, positions 19 to 454 of SEQ ID NO:22, positions 19 to 447 of SEQ ID NO:23, positions 19 to 447 of SEQ ID NO:24, positions 20 to 443 of SEQ ID NO:25, positions 18 to 447 of SEQ ID NO:26, positions 19 to 442 of SEQ ID NO:27, positions 18 to 451 of SEQ ID NO:28, positions 23 to 446 of SEQ ID NO:29, positions 18 to 444 of SEQ ID NO:30, positions 18 to 451 of SEQ ID NO:31, positions 18 to 447 of SEQ ID NO:32, positions 19 to 449 of SEQ ID NO:33, positions 18 to 447 of SEQ ID NO:34, positions 26 to 459 of SEQ ID NO:35, positions 19 to 450 of SEQ ID NO:36, positions 19 to 453 of SEQ ID NO:37, positions 18 to 448 of SEQ ID NO:38, positions 19 to 443 of SEQ ID NO:39, positions 19 to 442 of SEQ ID NO:40, positions 18 to 444 of SEQ ID NO:41, positions 24 to 457 of SEQ ID NO:42, positions 18 to 449 of SEQ ID NO:43, positions 19 to 453 of SEQ ID NO:44, positions 26 to 456 of SEQ ID NO:45, positions 19 to 451 of SEQ ID NO:46, positions 18 to 443 of SEQ ID NO:47, positions 18 to 448 of SEQ ID NO:48, positions 19 to 451 of SEQ ID NO:49, positions 18 to 444 of SEQ ID NO:50, positions 2 to 419 of SEQ ID NO:51, positions 27 to 461 of SEQ ID NO:52, positions 21 to 445 of SEQ ID NO:53, positions 19 to 449 of SEQ ID NO:54, positions 19 to 448 of SEQ ID NO:55, positions 18 to 443 of SEQ ID NO:56, positions 20 to 443 of SEQ ID NO:57, positions 18 to 448 of SEQ ID NO:58, positions 18 to 447 of SEQ ID NO:59, positions 26 to 455 of SEQ ID NO:60, positions 19 to 449 of SEQ ID NO:61, positions 19 to 449 of SEQ ID NO:62, positions 26 to 460 of SEQ ID NO:63, positions 18 to 448 of SEQ ID NO:64, positions 19 to 451 of SEQ ID NO:65, positions 19 to 447 of SEQ ID NO:66, positions 1 to 424 of SEQ ID NO:67, positions 19 to 448 of SEQ ID NO:68, positions 19 to 443 of SEQ ID NO:69, positions 23 to 447 of SEQ ID NO:70, positions 17 to 448 of SEQ ID NO:71, positions 19 to 449 of SEQ ID NO:72, positions 18 to 444 of SEQ ID NO:73, positions 23 to 458 of SEQ ID NO:74, positions 20 to 452 of SEQ ID NO:75, positions 18 to 435 of SEQ ID NO:76, positions 18 to 446 of SEQ ID NO:77, positions 22 to 457 of SEQ ID NO:78, positions 18 to 448 of SEQ ID NO:79, positions 1 to 431 of SEQ ID NO:80, positions 19 to 453 of SEQ ID NO:81, positions 21 to 440 of SEQ ID NO:82, positions 19 to 442 of SEQ ID NO:83, positions 18 to 448 of SEQ ID NO:84, positions 17 to 446 of SEQ ID NO:85, positions 18 to 447 of SEQ ID NO:86, positions 18 to 443 of SEQ ID NO:87, positions 23 to 448 of SEQ ID NO:88, positions 18 to 451 of SEQ ID NO:89, positions 21 to 447 of SEQ ID NO:90, positions 18 to 444 of SEQ ID NO:91, positions 19 to 442 of SEQ ID NO:92, positions 20 to 436 of SEQ ID NO:93, positions 18 to 450 of SEQ ID NO:94, positions 22 to 453 of SEQ ID NO:95, positions 16 to 472 of SEQ ID NO:96, positions 21 to 445 of SEQ ID NO:97, positions 19 to 447 of SEQ ID NO:98, positions 19 to 450 of SEQ ID NO:99, positions 19 to 451 of SEQ ID NO:100, positions 18 to 448 of SEQ ID NO:101, positions 19 to 442 of SEQ ID NO:102, positions 20 to 457 of SEQ ID NO:103, positions 19 to 454 of SEQ ID NO:104, positions 18 to 440 of SEQ ID NO:105, positions 18 to 439 of SEQ ID NO:106, positions 27 to 460 of SEQ ID NO:107, positions 23 to 446 of SEQ ID NO:108, positions 17 to 446 of SEQ ID NO:109, positions 21 to 447 of SEQ ID NO:110, positions 19 to 447 of SEQ ID NO:111, positions 18 to 449 of SEQ ID NO:112, positions 22 to 457 of SEQ ID NO:113, positions 18 to 445 of

SEQ ID NO:114, positions 18 to 448 of SEQ ID NO:115, positions 18 to 448 of SEQ ID NO:116, positions 23 to 435 of SEQ ID NO:117, positions 21 to 442 of SEQ ID NO:118, positions 23 to 435 of SEQ ID NO:119, positions 20 to 445 of SEQ ID NO:120, positions 21 to 443 of SEQ ID NO:121, positions 20 to 445 of SEQ ID NO:122, positions 23 to 443 of SEQ ID NO:123, positions 20 to 445 of SEQ ID NO:124, positions 21 to 435 of SEQ ID NO:125, positions 20 to 437 of SEQ ID NO:126, positions 21 to 442 of SEQ ID NO:127, positions 23 to 434 of SEQ ID NO:128, positions 20 to 444 of SEQ ID NO:129, positions 21 to 435 of SEQ ID NO:130, positions 20 to 445 of SEQ ID NO:131, positions 21 to 446 of SEQ ID NO:132, positions 21 to 435 of SEQ ID NO:133, positions 22 to 448 of SEQ ID NO:134, positions 23 to 433 of SEQ ID NO:135, positions 23 to 434 of SEQ ID NO:136, positions 23 to 435 of SEQ ID NO:137, positions 23 to 435 of SEQ ID NO:138, positions 20 to 445 of SEQ ID NO:139, positions 20 to 437 of SEQ ID NO:140, positions 21 to 435 of SEQ ID NO:141, positions 20 to 437 of SEQ ID NO:142, positions 21 to 435 of SEQ ID NO:143, positions 26 to 435 of SEQ ID NO:144, positions 23 to 435 of SEQ ID NO:145, positions 24 to 443 of SEQ ID NO:146, positions 20 to 445 of SEQ ID NO:147, positions 21 to 441 of SEQ ID NO:148, and positions 20 to 437 of SEQ ID NO:149.

The CBDs are particularly involved in the hydrolysis of crystalline cellulose. It has been shown that the ability of cellobiohydrolases to degrade crystalline cellulose decreases when the CBD is absent (Linder and Teeri, 1997, Journal of Biotechnol. 57:15-28). The variant CBH I polypeptides of the disclosure can further include a CBD. Exemplary CBDs comprise amino acid sequences corresponding to positions 494 to 529 of SEQ ID NO:1, positions 480 to 514 of SEQ ID NO:2, positions 494 to 529 of SEQ ID NO:3, positions 491 to 526 of SEQ ID NO:5, positions 477 to 512 of SEQ ID NO:6, positions 497 to 532 of SEQ ID NO:7, positions 504 to 539 of SEQ ID NO:8, positions 486 to 521 of SEQ ID NO:13, positions 556 to 596 of SEQ ID NO:15, positions 490 to 525 of SEQ ID NO:18, positions 495 to 530 of SEQ ID NO:20, positions 471 to 506 of SEQ ID NO:23, positions 481 to 516 of SEQ ID NO:27, positions 480 to 514 of SEQ ID NO:30, positions 495 to 529 of SEQ ID NO:35, positions 493 to 528 of SEQ ID NO:36, positions 477 to 512 of SEQ ID NO:38, positions 547 to 586 of SEQ ID NO:39, positions 475 to 510 of SEQ ID NO:40, positions 479 to 513 of SEQ ID NO:41, positions 506 to 541 of SEQ ID NO:42, positions 481 to 516 of SEQ ID NO:43, positions 503 to 537 of SEQ ID NO:45, positions 488 to 523 of SEQ ID NO:46, positions 476 to 511 of SEQ ID NO:48, positions 488 to 523 of SEQ ID NO:49, positions 479 to 513 of SEQ ID NO:50, positions 500 to 535 of SEQ ID NO:52, positions 493 to 528 of SEQ ID NO:55, positions 479 to 514 of SEQ ID NO:58, positions 494 to 529 of SEQ ID NO:60, positions 490 to 525 of SEQ ID NO:61, positions 497 to 532 of SEQ ID NO:62, positions 475 to 510 of SEQ ID NO:64, positions 477 to 512 of SEQ ID NO:65, positions 486 to 521 of SEQ ID NO:66, positions 470 to 505 of SEQ ID NO:67, positions 491 to 526 of SEQ ID NO:68, positions 476 to 511 of SEQ ID NO:69, positions 480 to 514 of SEQ ID NO:73, positions 506 to 540 of SEQ ID NO:74, positions 471 to 504 of SEQ ID NO:76, positions 501 to 536 of SEQ ID NO:78, positions 473 to 508 of SEQ ID NO:79, positions 481 to 516 of SEQ ID NO:83, positions 488 to 523 of SEQ ID NO:86, positions 475 to 510 of SEQ ID NO:92, positions 468 to 504 of SEQ ID NO:93, positions 501 to 536 of SEQ ID NO:96, positions 482 to 517 of SEQ ID NO:98, positions 481 to 516 of SEQ ID NO:99, positions 488 to 523 of SEQ ID NO:100, positions 472 to 507 of SEQ ID NO:101, positions 481 to 516 of SEQ ID NO:102, positions 471 to 505

of SEQ ID NO:105, positions 481 to 516 of SEQ ID NO:106, positions 495 to 530 of SEQ ID NO:107, positions 488 to 523 of SEQ ID NO:111, positions 478 to 513 of SEQ ID NO:112, positions 501 to 536 of SEQ ID NO:113, positions 491 to 526 of SEQ ID NO:115, and positions 503 to 538 of SEQ ID NO:116.

The CD and CBD are often connected via a linker. Exemplary linker sequences correspond to positions 456 to 493 of SEQ ID NO:1, positions 445 to 479 of SEQ ID NO:2, positions 456 to 493 of SEQ ID NO:3, positions 458 to 490 of SEQ ID NO:5, positions 449 to 476 of SEQ ID NO:6, positions 461 to 496 of SEQ ID NO:7, positions 461 to 503 of SEQ ID NO:8, positions 446 to 485 of SEQ ID NO:13, positions 444 to 555 of SEQ ID NO:15, positions 450 to 489 of SEQ ID NO:18, positions 450 to 494 of SEQ ID NO:20, positions 448 to 470 of SEQ ID NO:23, positions 443 to 480 of SEQ ID NO:27, positions 445 to 479 of SEQ ID NO:30, positions 460 to 494 of SEQ ID NO:35, positions 451 to 492 of SEQ ID NO:36, positions 449 to 476 of SEQ ID NO:38, positions 444 to 546 of SEQ ID NO:39, positions 443 to 474 of SEQ ID NO:40, positions 445 to 478 of SEQ ID NO:41, positions 458 to 505 of SEQ ID NO:42, positions 450 to 480 of SEQ ID NO:43, positions 457 to 502 of SEQ ID NO:45, positions 452 to 487 of SEQ ID NO:46, positions 449 to 475 of SEQ ID NO:48, positions 452 to 487 of SEQ ID NO:49, positions 445 to 478 of SEQ ID NO:50, positions 462 to 499 of SEQ ID NO:52, positions 449 to 492 of SEQ ID NO:55, positions 449 to 478 of SEQ ID NO:58, positions 456 to 493 of SEQ ID NO:60, positions 450 to 489 of SEQ ID NO:61, positions 450 to 496 of SEQ ID NO:62, positions 449 to 474 of SEQ ID NO:64, positions 452 to 476 of SEQ ID NO:65, positions 448 to 485 of SEQ ID NO:66, positions 425 to 469 of SEQ ID NO:67, positions 449 to 490 of SEQ ID NO:68, positions 444 to 475 of SEQ ID NO:69, positions 445 to 479 of SEQ ID NO:73, positions 459 to 505 of SEQ ID NO:74, positions 436 to 470 of SEQ ID NO:76, positions 458 to 500 of SEQ ID NO:78, positions 449 to 472 of SEQ ID NO:79, positions 443 to 480 of SEQ ID NO:83, positions 448 to 487 of SEQ ID NO:86, positions 443 to 474 of SEQ ID NO:92, positions 437 to 467 of SEQ ID NO:93, positions 473 to 500 of SEQ ID NO:96, positions 448 to 481 of SEQ ID NO:98, positions 451 to 480 of SEQ ID NO:99, positions 452 to 487 of SEQ ID NO:100, positions 449 to 471 of SEQ ID NO:101, positions 443 to 480 of SEQ ID NO:102, positions 441 to 470 of SEQ ID NO:105, positions 440 to 480 of SEQ ID NO:106, positions 461 to 494 of SEQ ID NO:107, positions 448 to 487 of SEQ ID NO:111, positions 450 to 478 of SEQ ID NO:112, positions 458 to 500 of SEQ ID NO:113, positions 449 to 490 of SEQ ID NO:115, and positions 449 to 502 of SEQ ID NO:116.

Because CBH I polypeptides are modular, the CBDs, CDs and linkers of different CBH I polypeptides, such as the exemplary CBH I polypeptides of Table 1, can be used interchangeably. However, in a preferred embodiment, the CBDs, CDs and linkers of a variant CBH I of the disclosure originate from the same polypeptide.

The variant CBH I polypeptides of the disclosure preferably have at least a two-fold reduction of product inhibition, such that cellobiose has an IC<sub>50</sub> towards the variant CBH I that is at least 2-fold the IC<sub>50</sub> of the corresponding reference CBH I, e.g., CBH I lacking the R268 substitution and/or R411 substitution. More preferably the IC<sub>50</sub> of cellobiose towards the variant CBH I is at least 3-fold, at least 5-fold, at least 8-fold, at least 10-fold, at least 12-fold or at least 15-fold the IC<sub>50</sub> of the corresponding reference CBH I. In specific embodiments the IC<sub>50</sub> of cellobiose towards the variant CBH I is ranges from 2-fold to 15-fold, from 2-fold to 10-fold, from

3-fold to 10-fold, from 5-fold to 12-fold, from 4-fold to 12-fold, from 5-fold to 10-fold, from 5-fold to 12-fold, from 2-fold to 8-fold, or from 8-fold to 20-fold the IC<sub>50</sub> of the corresponding reference CBH I. The IC<sub>50</sub> can be determined in a phosphoric acid swollen cellulose ("PASC") assay (Du et al., 2010, Applied Biochemistry and Biotechnology 161:313-317) or a methylumbelliferyl lactoside ("MUL") assay (van Tilburgh and Claeyssens, 1985, FEBS Letts. 187(2):283-288), as exemplified in the Examples below.

The variant CBH I polypeptides of the disclosure preferably have a cellobiohydrolase activity that is at least 30% the cellobiohydrolase activity of the corresponding reference CBH I, e.g., CBH I lacking the R268 substitution and/or R411 substitution. More preferably, the cellobiohydrolase activity of the variant CBH I is at least 40%, at least 50%, at least 60% or at least 70% the cellobiohydrolase activity of the corresponding reference CBH I. In specific embodiments the IC<sub>50</sub> cellobiohydrolase activity of the variant CBH I is ranges from 30% to 80%, from 40% to 70%, 30% to 60%, from 50% to 80% or from 60% to 80% of the cellobiohydrolase activity of the corresponding reference CBH I. Assays for cellobiohydrolase activity are described, for example, in Becker et al., 2011, Biochem J. 356:19-30 and Mitsuishi et al., 1990, FEBS Letts. 275:135-138, each of which is expressly incorporated by reference herein. The ability of CBH I to hydrolyze isolated soluble and insoluble substrates can also be measured using assays described in Srisodsuk et al., 1997, J. Biotech. 57:4957 and Nidetzky and Claeyssens, 1994, Biotech. Bioeng. 44:961-966. Substrates useful for assaying cellobiohydrolase activity include crystalline cellulose, filter paper, phosphoric acid swollen cellulose, cellooligosaccharides, methylumbelliferyl lactoside, methylumbelliferyl cellobioside, orthonitrophenyl lactoside, paranitrophenyl lactoside, orthorotophenyl cellobioside, paranitrophenyl cellobioside. Cellobiohydrolase activity can be measured in an assay utilizing PASC as the substrate and a calcofluor white detection method (Du et al., 2010, Applied Biochemistry and Biotechnology 161:313-317). PASC can be prepared as described by Walseth, 1952, TAPPI 35:228-235 and Wood, 1971, Biochem. J. 121:353-362.

Other than said R268 and/or R411 substitution, the variant CBH I polypeptides of the disclosure preferably:

comprise an amino acid sequence having at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity to a CD of a reference CBH I exemplified in Table 1 (i.e., a CD comprising an amino acid sequence corresponding to positions 26 to 455 of SEQ ID NO:1, positions 18 to 444 of SEQ ID NO:2, positions 26 to 455 of SEQ ID NO:3, positions 1 to 427 of SEQ ID NO:4, positions 24 to 457 of SEQ ID NO:5, positions 18 to 448 of SEQ ID NO:6, positions 27 to 460 of SEQ ID NO:7, positions 27 to 460 of SEQ ID NO:8, positions 20 to 449 of SEQ ID NO:9, positions 1 to 424 of SEQ ID NO:10, positions 18 to 447 of SEQ ID NO:11, positions 18 to 434 of SEQ ID NO:12, positions 18 to 445 of SEQ ID NO:13, positions 19 to 454 of SEQ ID NO:14, positions 19 to 443 of SEQ ID NO:15, positions 2 to 426 of SEQ ID NO:16, positions 23 to 446 of SEQ ID NO:17, positions 19 to 449 of SEQ ID NO:18, positions 23 to 446 of SEQ ID NO:19, positions 19 to 449 of SEQ ID NO:20, positions 2 to 416 of SEQ ID NO:21, positions 19 to 454 of SEQ ID NO:22, positions 19 to 447 of SEQ

ID NO:23, positions 19 to 447 of SEQ ID NO:24, positions 20 to 443 of SEQ ID NO:25, positions 18 to 447 of SEQ ID NO:26, positions 19 to 442 of SEQ ID NO:27, positions 18 to 451 of SEQ ID NO:28, positions 23 to 446 of SEQ ID NO:29, positions 18 to 444 of SEQ ID NO:30, positions 18 to 451 of SEQ ID NO:31, positions 18 to 447 of SEQ ID NO:32, positions 19 to 449 of SEQ ID NO:33, positions 18 to 447 of SEQ ID NO:34, positions 26 to 459 of SEQ ID NO:35, positions 19 to 450 of SEQ ID NO:36, positions 19 to 453 of SEQ ID NO:37, positions 18 to 448 of SEQ ID NO:38, positions 19 to 443 of SEQ ID NO:39, positions 19 to 442 of SEQ ID NO:40, positions 18 to 444 of SEQ ID NO:41, positions 24 to 457 of SEQ ID NO:42, positions 18 to 449 of SEQ ID NO:43, positions 19 to 453 of SEQ ID NO:44, positions 26 to 456 of SEQ ID NO:45, positions 19 to 451 of SEQ ID NO:46, positions 18 to 443 of SEQ ID NO:47, positions 18 to 448 of SEQ ID NO:48, positions 19 to 451 of SEQ ID NO:49, positions 18 to 444 of SEQ ID NO:50, positions 2 to 419 of SEQ ID NO:51, positions 27 to 461 of SEQ ID NO:52, positions 21 to 445 of SEQ ID NO:53, positions 19 to 449 of SEQ ID NO:54, positions 19 to 448 of SEQ ID NO:55, positions 18 to 443 of SEQ ID NO:56, positions 20 to 443 of SEQ ID NO:57, positions 18 to 448 of SEQ ID NO:58, positions 18 to 447 of SEQ ID NO:59, positions 26 to 455 of SEQ ID NO:60, positions 19 to 449 of SEQ ID NO:61, positions 19 to 449 of SEQ ID NO:62, positions 26 to 460 of SEQ ID NO:63, positions 18 to 448 of SEQ ID NO:64, positions 19 to 451 of SEQ ID NO:65, positions 19 to 447 of SEQ ID NO:66, positions 1 to 424 of SEQ ID NO:67, positions 19 to 448 of SEQ ID NO:68, positions 19 to 443 of SEQ ID NO:69, positions 23 to 447 of SEQ ID NO:70, positions 17 to 448 of SEQ ID NO:71, positions 19 to 449 of SEQ ID NO:72, positions 18 to 444 of SEQ ID NO:73, positions 23 to 458 of SEQ ID NO:74, positions 20 to 452 of SEQ ID NO:75, positions 18 to 435 of SEQ ID NO:76, positions 18 to 446 of SEQ ID NO:77, positions 22 to 457 of SEQ ID NO:78, positions 18 to 448 of SEQ ID NO:79, positions 1 to 431 of SEQ ID NO:80, positions 19 to 453 of SEQ ID NO:81, positions 21 to 440 of SEQ ID NO:82, positions 19 to 442 of SEQ ID NO:83, positions 18 to 448 of SEQ ID NO:84, positions 17 to 446 of SEQ ID NO:85, positions 18 to 447 of SEQ ID NO:86, positions 18 to 443 of SEQ ID NO:87, positions 23 to 448 of SEQ ID NO:88, positions 18 to 451 of SEQ ID NO:89, positions 21 to 447 of SEQ ID NO:90, positions 18 to 444 of SEQ ID NO:91, positions 19 to 442 of SEQ ID NO:92, positions 20 to 436 of SEQ ID NO:93, positions 18 to 450 of SEQ ID NO:94, positions 22 to 453 of SEQ ID NO:95, positions 16 to 472 of SEQ ID NO:96, positions 21 to 445 of SEQ ID NO:97, positions 19 to 447 of SEQ ID NO:98, positions 19 to 450 of SEQ ID NO:99, positions 19 to 451 of SEQ ID NO:100, positions 18 to 448 of SEQ ID NO:101, positions 19 to 442 of SEQ ID NO:102, positions 20 to 457 of SEQ ID NO:103, positions 19 to 454 of SEQ ID NO:104, positions 18 to 440 of SEQ ID NO:105, positions 18 to 439 of SEQ ID NO:106, positions 27 to 460 of SEQ ID NO:107, positions 23 to 446 of SEQ ID NO:108, positions 17 to 446 of SEQ ID NO:109, positions 21 to 447 of SEQ ID NO:110, positions 19 to 447 of SEQ ID NO:111, positions 18 to 449 of SEQ ID NO:112, positions 22 to 457 of SEQ ID NO:113, positions 18 to 445 of SEQ ID NO:114, positions 18 to 448 of SEQ ID NO:115, positions 18 to 448 of SEQ ID NO:116, positions 23 to 435 of SEQ ID NO:117, posi-

**11**

tions 21 to 442 of SEQ ID NO:118, positions 23 to 435 of SEQ ID NO:119, positions 20 to 445 of SEQ ID NO:120, positions 21 to 443 of SEQ ID NO:121, positions 20 to 445 of SEQ ID NO:122, positions 23 to 443 of SEQ ID NO:123, positions 20 to 445 of SEQ ID NO:124, positions 21 to 435 of SEQ ID NO:125, positions 20 to 437 of SEQ ID NO:126, positions 21 to 442 of SEQ ID NO:127, positions 23 to 434 of SEQ ID NO:128, positions 20 to 444 of SEQ ID NO:129, positions 21 to 435 of SEQ ID NO:130, positions 20 to 445 of SEQ ID NO:131, positions 21 to 446 of SEQ ID NO:132, positions 21 to 435 of SEQ ID NO:133, positions 22 to 448 of SEQ ID NO:134, positions 23 to 433 of SEQ ID NO:135, positions 23 to 434 of SEQ ID NO:136, positions 23 to 435 of SEQ ID NO:137, positions 23 to 435 of SEQ ID NO:138, positions 20 to 445 of SEQ ID NO:139, positions 20 to 437 of SEQ ID NO:140, positions 21 to 435 of SEQ ID NO:141, positions 20 to 437 of SEQ ID NO:142, positions 21 to 435 of SEQ ID NO:143, positions 26 to 435 of SEQ ID NO:144, positions 23 to 435 of SEQ ID NO:145, positions 24 to 443 of SEQ ID NO:146, positions 20 to 445 of SEQ ID NO:147, positions 21 to 441 of SEQ ID NO:148, and positions 20 to 437 of SEQ ID NO:149 (preferably the CD corresponding to positions 26-455 of SEQ ID NO: 1 or 18-444 of SEQ ID NO:2); and/or comprise an amino acid sequence having at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity to a mature polypeptide of a reference CBH I exemplified in Table 1 (i.e., a mature protein comprising an amino acid sequence corresponding to positions 26 to 529 of SEQ ID NO:1, positions 18 to 514 of SEQ ID NO:2, positions 26 to 529 of SEQ ID NO:3, positions 1 to 427 of SEQ ID NO:4, positions 24 to 526 of SEQ ID NO:5, positions 18 to 512 of SEQ ID NO:6, positions 27 to 532 of SEQ ID NO:7, positions 27 to 539 of SEQ ID NO:8, positions 20 to 449 of SEQ ID NO:9, positions 1 to 424 of SEQ ID NO:10, positions 18 to 447 of SEQ ID NO:11, positions 18 to 434 of SEQ ID NO:12, positions 18 to 521 of SEQ ID NO:13, positions 19 to 454 of SEQ ID NO:14, positions 19 to 596 of SEQ ID NO:15, positions 2 to 426 of SEQ ID NO:16, positions 23 to 446 of SEQ ID NO:17, positions 19 to 525 of SEQ ID NO:18, positions 23 to 446 of SEQ ID NO:19, positions 19 to 530 of SEQ ID NO:20, positions 2 to 416 of SEQ ID NO:21, positions 19 to 454 of SEQ ID NO:22, positions 19 to 506 of SEQ ID NO:23, positions 19 to 447 of SEQ ID NO:24, positions 20 to 443 of SEQ ID NO:25, positions 18 to 447 of SEQ ID NO:26, positions 19 to 516 of SEQ ID NO:27, positions 18 to 451 of SEQ ID NO:28, positions 23 to 446 of SEQ ID NO:29, positions 18 to 514 of SEQ ID NO:30, positions 18 to 451 of SEQ ID NO:31, positions 18 to 447 of SEQ ID NO:32, positions 19 to 449 of SEQ ID NO:33, positions 18 to 447 of SEQ ID NO:34, positions 26 to 529 of SEQ ID NO:35, positions 19 to 528 of SEQ ID NO:36, positions 19 to 453 of SEQ ID NO:37, positions 18 to 512 of SEQ ID NO:38, positions 19 to 586 of SEQ ID NO:39, positions 19 to 510 of SEQ ID NO:40, positions 18 to 513 of SEQ ID NO:41, positions 24 to 541 of SEQ ID NO:42, positions 18 to 516 of SEQ ID NO:43, positions 19 to 453 of SEQ ID NO:44, positions 26 to 537 of SEQ

**12**

ID NO:45, positions 19 to 523 of SEQ ID NO:46, positions 18 to 443 of SEQ ID NO:47, positions 18 to 511 of SEQ ID NO:48, positions 19 to 523 of SEQ ID NO:49, positions 18 to 513 of SEQ ID NO:50, positions 2 to 419 of SEQ ID NO:51, positions 27 to 535 of SEQ ID NO:52, positions 21 to 445 of SEQ ID NO:53, positions 19 to 449 of SEQ ID NO:54, positions 19 to 528 of SEQ ID NO:55, positions 18 to 443 of SEQ ID NO:56, positions 20 to 443 of SEQ ID NO:57, positions 18 to 514 of SEQ ID NO:58, positions 18 to 447 of SEQ ID NO:59, positions 26 to 529 of SEQ ID NO:60, positions 19 to 525 of SEQ ID NO:61, positions 19 to 532 of SEQ ID NO:62, positions 26 to 460 of SEQ ID NO:63, positions 18 to 510 of SEQ ID NO:64, positions 19 to 512 of SEQ ID NO:65, positions 19 to 521 of SEQ ID NO:66, positions 1 to 505 of SEQ ID NO:67, positions 19 to 526 of SEQ ID NO:68, positions 19 to 511 of SEQ ID NO:69, positions 23 to 447 of SEQ ID NO:70, positions 17 to 448 of SEQ ID NO:71, positions 19 to 449 of SEQ ID NO:72, positions 18 to 514 of SEQ ID NO:73, positions 23 to 540 of SEQ ID NO:74, positions 20 to 452 of SEQ ID NO:75, positions 18 to 504 of SEQ ID NO:76, positions 18 to 446 of SEQ ID NO:77, positions 22 to 536 of SEQ ID NO:78, positions 18 to 508 of SEQ ID NO:79, positions 1 to 431 of SEQ ID NO:80, positions 19 to 453 of SEQ ID NO:81, positions 21 to 440 of SEQ ID NO:82, positions 19 to 516 of SEQ ID NO:83, positions 18 to 448 of SEQ ID NO:84, positions 17 to 446 of SEQ ID NO:85, positions 18 to 523 of SEQ ID NO:86, positions 18 to 443 of SEQ ID NO:87, positions 23 to 448 of SEQ ID NO:88, positions 18 to 451 of SEQ ID NO:89, positions 21 to 447 of SEQ ID NO:90, positions 18 to 444 of SEQ ID NO:91, positions 19 to 510 of SEQ ID NO:92, positions 20 to 504 of SEQ ID NO:93, positions 18 to 450 of SEQ ID NO:94, positions 22 to 453 of SEQ ID NO:95, positions 16 to 536 of SEQ ID NO:96, positions 21 to 445 of SEQ ID NO:97, positions 19 to 517 of SEQ ID NO:98, positions 19 to 516 of SEQ ID NO:99, positions 19 to 523 of SEQ ID NO:100, positions 18 to 507 of SEQ ID NO:101, positions 19 to 516 of SEQ ID NO:102, positions 20 to 457 of SEQ ID NO:103, positions 19 to 454 of SEQ ID NO:104, positions 18 to 505 of SEQ ID NO:105, positions 18 to 516 of SEQ ID NO:106, positions 27 to 530 of SEQ ID NO:107, positions 23 to 446 of SEQ ID NO:108, positions 17 to 446 of SEQ ID NO:109, positions 21 to 447 of SEQ ID NO:110, positions 19 to 523 of SEQ ID NO:111, positions 18 to 513 of SEQ ID NO:112, positions 22 to 536 of SEQ ID NO:113, positions 18 to 445 of SEQ ID NO:114, positions 18 to 526 of SEQ ID NO:115, positions 18 to 538 of SEQ ID NO:116, positions 23 to 435 of SEQ ID NO:117, positions 21 to 442 of SEQ ID NO:118, positions 23 to 435 of SEQ ID NO:119, positions 20 to 445 of SEQ ID NO:120, positions 21 to 443 of SEQ ID NO:121, positions 20 to 445 of SEQ ID NO:122, positions 23 to 443 of SEQ ID NO:123, positions 20 to 445 of SEQ ID NO:124, positions 21 to 435 of SEQ ID NO:125, positions 20 to 437 of SEQ ID NO:126, positions 21 to 442 of SEQ ID NO:127, positions 23 to 434 of SEQ ID NO:128, positions 20 to 444 of SEQ ID NO:129, positions 21 to 435 of SEQ ID NO:130, positions 20 to 445 of SEQ ID NO:131, positions 21 to 446 of SEQ ID NO:132, positions 21 to 435 of SEQ ID NO:133, positions 22 to 448 of SEQ ID NO:134, positions 23 to 433 of SEQ ID NO:135, positions 23 to 434 of SEQ ID NO:136, positions 23 to 435 of SEQ ID NO:137, positions 23 to 435 of SEQ ID

13

NO:138, positions 20 to 445, of SEQ ID NO:139, positions 20 to 437 of SEQ ID NO:140, positions 21 to 435 of SEQ ID NO:141, positions 20 to 437 of SEQ ID NO:142, positions 21 to 435 of SEQ ID NO:143, positions 26 to 435 of SEQ ID NO:144, positions 23 to 435 of SEQ ID NO:145, positions 24 to 443 of SEQ ID NO:146, positions 20 to 445 of SEQ ID NO:147, positions 21 to 441 of SEQ ID NO:148, and positions 20 to 437 of SEQ ID NO:149, preferably the mature polypeptide corresponding to positions 26-529 of SEQ ID NO:1 or 18-514 of SEQ ID NO:2.

An example of an algorithm that is suitable for determining sequence similarity is the BLAST algorithm, which is described in Altschul et al., 1990, *J. Mol. Biol.* 215:403-410. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. These initial neighborhood word hits act as starting points to find longer HSPs containing them. The word hits are expanded in both directions along each of the two sequences being compared for as far as the cumulative alignment score can be increased. Extension of the word hits is stopped when: the cumulative alignment score falls off by the quantity X from a maximum achieved value; the cumulative score goes to zero or below; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a word length (W) of 11, the BLOSUM62 scoring matrix (see Henikoff & Henikoff, 1992, *Proc. Nat'l. Acad. Sci. USA* 89:10915-10919) alignments (B) of 50, expectation (E) of 10, M'S, N'-4, and a comparison of both strands.

Most CBH I polypeptides are secreted and are therefore expressed with a signal sequence that is cleaved upon secretion of the polypeptide from the cell. Accordingly, in certain aspects, the variant CBH I polypeptides of the disclosure further include a signal sequence. Exemplary signal sequences comprise amino acid sequences corresponding to positions 1 to 25 of SEQ ID NO:1, positions 1 to 17 of SEQ ID NO:2, positions 1 to 25 of SEQ ID NO:3, positions 1 to 23 of SEQ ID NO:5, positions 1 to 17 of SEQ ID NO:6, positions 1 to 26 of SEQ ID NO:7, positions 1 to 27 of SEQ ID NO:8, positions 1 to 19 of SEQ ID NO:9, positions 1 to 17 of SEQ ID NO:11, positions 1 to 17 of SEQ ID NO:12, positions 1 to 17 of SEQ ID NO:13, positions 1 to 18 of SEQ ID NO:14, positions 1 to 18 of SEQ ID NO:15, positions 1 to 22 of SEQ ID NO:17, positions 1 to 18 of SEQ ID NO:18, positions 1 to 22 of SEQ ID NO:19, positions 1 to 18 of SEQ ID NO:20, positions 1 to 18 of SEQ ID NO:22, positions 1 to 18 of SEQ ID NO:23, positions 1 to 18 of SEQ ID NO:24, positions 1 to 19 of SEQ ID NO:25, positions 1 to 17 of SEQ ID NO:26, positions 1 to 18 of SEQ ID NO:27, positions 1 to 17 of SEQ ID NO:28, positions 1 to 22 of SEQ ID NO:29, positions 1 to 18 of SEQ ID NO:30, positions 1 to 17 of SEQ ID NO:31, positions 1 to 17 of SEQ ID NO:32, positions 1 to 18 of SEQ ID NO:33, positions 1 to 17 of SEQ ID NO:34, positions 1 to 25 of SEQ ID NO:35, positions 1 to 18 of SEQ ID NO:36, positions 1 to 18 of SEQ ID NO:37, positions 1 to 17 of SEQ ID NO:38, positions 1 to 18 of SEQ ID NO:39, positions 1 to 18 of SEQ ID NO:40, positions 1 to 17 of SEQ ID NO:41, positions 1 to 23 of SEQ ID NO:42, positions 1 to 17 of SEQ ID NO:43, positions 1 to 18 of SEQ ID NO:44, positions 1 to 25 of SEQ ID NO:45, positions 1 to 18 of SEQ ID NO:46, positions 1 to 17 of SEQ ID NO:47, positions 1 to 17 of SEQ

14

ID NO:48, positions 1 to 18 of SEQ ID NO:49, positions 1 to 17 of SEQ ID NO:50, positions 1 to 26 of SEQ ID NO:52, positions 1 to 20 of SEQ ID NO:53, positions 1 to 18 of SEQ ID NO:54, positions 1 to 18 of SEQ ID NO:55, positions 1 to 17 of SEQ ID NO:56, positions 1 to 19 of SEQ ID NO:57, positions 1 to 17 of SEQ ID NO:58, positions 1 to 17 of SEQ ID NO:59, positions 1 to 25 of SEQ ID NO:60, positions 1 to 18 of SEQ ID NO:61, positions 1 to 18 of SEQ ID NO:62, positions 1 to 25 of SEQ ID NO:63, positions 1 to 17 of SEQ ID NO:64, positions 1 to 18 of SEQ ID NO:65, positions 1 to 18 of SEQ ID NO:66, positions 1 to 18 of SEQ ID NO:68, positions 1 to 18 of SEQ ID NO:69, positions 1 to 23 of SEQ ID NO:70, positions 1 to 17 of SEQ ID NO:71, positions 1 to 18 of SEQ ID NO:72, positions 1 to 17 of SEQ ID NO:73, positions 1 to 22 of SEQ ID NO:74, positions 1 to 19 of SEQ ID NO:75, positions 1 to 17 of SEQ ID NO:76, positions 1 to 17 of SEQ ID NO:77, positions 1 to 21 of SEQ ID NO:78, positions 1 to 18 of SEQ ID NO:79, positions 1 to 18 of SEQ ID NO:81, positions 1 to 20 of SEQ ID NO:82, positions 1 to 18 of SEQ ID NO:83, positions 1 to 17 of SEQ ID NO:84, positions 1 to 16 of SEQ ID NO:85, positions 1 to 17 of SEQ ID NO:86, positions 1 to 17 of SEQ ID NO:87, positions 1 to 22 of SEQ ID NO:88, positions 1 to 17 of SEQ ID NO:89, positions 1 to 20 of SEQ ID NO:90, positions 1 to 17 of SEQ ID NO:91, positions 1 to 18 of SEQ ID NO:92, positions 1 to 19 of SEQ ID NO:93, positions 1 to 17 of SEQ ID NO:94, positions 1 to 21 of SEQ ID NO:95, positions 1 to 15 of SEQ ID NO:96, positions 1 to 20 of SEQ ID NO:97, positions 1 to 18 of SEQ ID NO:98, positions 1 to 18 of SEQ ID NO:99, positions 1 to 18 of SEQ ID NO:100, positions 1 to 17 of SEQ ID NO:101, positions 1 to 18 of SEQ ID NO:102, positions 1 to 19 of SEQ ID NO:103, positions 1 to 18 of SEQ ID NO:104, positions 1 to 17 of SEQ ID NO:105, positions 1 to 17 of SEQ ID NO:106, positions 1 to 26 of SEQ ID NO:107, positions 1 to 22 of SEQ ID NO:108, positions 1 to 16 of SEQ ID NO:109, positions 1 to 20 of SEQ ID NO:110, positions 1 to 18 of SEQ ID NO:111, positions 1 to 17 of SEQ ID NO:112, positions 1 to 21 of SEQ ID NO:113, positions 1 to 17 of SEQ ID NO:114, positions 1 to 17 of SEQ ID NO:115, positions 1 to 18 of SEQ ID NO:116, positions 1 to 22 of SEQ ID NO:117, positions 1 to 20 of SEQ ID NO:118, positions 1 to 22 of SEQ ID NO:119, positions 1 to 19 of SEQ ID NO:120, positions 1 to 20 of SEQ ID NO:121, positions 1 to 19 of SEQ ID NO:122, positions 1 to 22 of SEQ ID NO:123, positions 1 to 19 of SEQ ID NO:124, positions 1 to 20 of SEQ ID NO:125, positions 1 to 19 of SEQ ID NO:126, positions 1 to 21 of SEQ ID NO:127, positions 1 to 22 of SEQ ID NO:128, positions 1 to 19 of SEQ ID NO:129, positions 1 to 20 of SEQ ID NO:130, positions 1 to 19 of SEQ ID NO:131, positions 1 to 20 of SEQ ID NO:132, positions 1 to 20 of SEQ ID NO:133, positions 1 to 21 of SEQ ID NO:134, positions 1 to 22 of SEQ ID NO:135, positions 1 to 22 of SEQ ID NO:136, positions 1 to 22 of SEQ ID NO:137, positions 1 to 22 of SEQ ID NO:138, positions 1 to 19 of SEQ ID NO:139, positions 1 to 19 of SEQ ID NO:140, positions 1 to 20 of SEQ ID NO:141, positions 1 to 19 of SEQ ID NO:142, positions 1 to 20 of SEQ ID NO:143, positions 1 to 25 of SEQ ID NO:144, positions 1 to 22 of SEQ ID NO:145, positions 1 to 23 of SEQ ID NO:146, positions 1 to 19 of SEQ ID NO:147, positions 1 to 20 of SEQ ID NO:148, and positions 1 to 19 of SEQ ID NO:149.

Recombinant Expression of Variant CBH I Polypeptides  
Cell Culture Systems

The disclosure also provides recombinant cells engineered to express variant CBH I polypeptides. Suitably, the variant CBH I polypeptide is encoded by a nucleic acid operably linked to a promoter.

Where recombinant expression in a filamentous fungal host is desired, the promoter can be a filamentous fungal promoter. The nucleic acids can be, for example, under the control of heterologous promoters. The variant CBH I polypeptides can also be expressed under the control of constitutive or inducible promoters. Examples of promoters that can be used include, but are not limited to, a cellulase promoter, a xylanase promoter, the 1818 promoter (previously identified as a highly expressed protein by EST mapping *Trichoderma*). For example, the promoter can suitably be a cellobiohydrolase, endoglucanase, or β-glucosidase promoter. A particularly suitable promoter can be, for example, a *T. reesei* cellobiohydrolase, endoglucanase, or β-glucosidase promoter. Non-limiting examples of promoters include a cbh1, cbh2, egl1, egl2, egl3, egl4, egl5, pki1, gpd1, xyn1, or xyn2 promoter.

Suitable host cells include cells of any microorganism (e.g., cells of a bacterium, a protist, an alga, a fungus (e.g., a yeast or filamentous fungus), or other microbe), and are preferably cells of a bacterium, a yeast, or a filamentous fungus.

Suitable host cells of the bacterial genera include, but are not limited to, cells of *Escherichia*, *Bacillus*, *Lactobacillus*, *Pseudomonas*, and *Streptomyces*. Suitable cells of bacterial species include, but are not limited to, cells of *Escherichia coli*, *Bacillus subtilis*, *Bacillus licheniformis*, *Lactobacillus brevis*, *Pseudomonas aeruginosa*, and *Streptomyces lividans*.

Suitable host cells of the genera of yeast include, but are not limited to, cells of *Saccharomyces*, *Schizosaccharomyces*, *Candida*, *Hansenula*, *Pichia*, *Kluyveromyces*, and *Phaffia*. Suitable cells of yeast species include, but are not limited to, cells of *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Candida albicans*, *Hansenula polymorpha*, *Pichia pastoris*, *P. canadensis*, *Kluyveromyces marxianus*, and *Phaf-fia rhodozyma*.

Suitable host cells of filamentous fungi include all filamentous forms of the subdivision Eumycotina. Suitable cells of filamentous fungal genera include, but are not limited to, cells of *Acremonium*, *Aspergillus*, *Aureobasidium*, *Bjerkandera*, *Ceriporiopsis*, *Chrysoporium*, *Coprinus*, *Coriolus*, *Corynascus*, *Chaetomium*, *Cryptococcus*, *Filobasidium*, *Fusarium*, *Gibberella*, *Humicola*, *Hypocrea*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Mucor*, *Neocallimastix*, *Neospora*, *Paecilomyces*, *Penicillium*, *Phanerochaete*, *Phlebia*, *Promyces*, *Pleurotus*, *Scytalidium*, *Schizophyllum*, *Sporotrichum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolyphocladium*, *Trametes*, and *Trichoderma*. More preferably, the recombinant cell is a *Trichoderma* sp. (e.g., *Trichoderma reesei*), *Penicillium* sp., *Humicola* sp. (e.g., *Humicola insolens*); *Aspergillus* sp. (e.g., *Aspergillus niger*), *Chrysosporium* sp., *Fusarium* sp., or *Hypocrea* sp. Suitable cells can also include cells of various anamorph and teleomorph forms of these filamentous fungal genera.

Suitable cells of filamentous fungal species include, but are not limited to, cells of *Aspergillus awamori*, *Aspergillus fumigatus*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Chrysosporium lucknowense*, *Fusarium bactridiooides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium gramininum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioïdes*, *Fusarium venenatum*, *Bjerkandera adusta*, *Ceriporiopsis aneirina*, *Ceriporiopsis aneirina*, *Ceriporiopsis caregiea*, *Ceriporiopsis gilvescens*, *Ceriporiopsis pannocinta*, *Ceriporiopsis rivulosa*, *Ceriporiopsis subrufa*, *Cerip-*

*riopsis subvermispora*, *Coprinus cinereus*, *Coriolus hirsutus*, *Humicola insolens*, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Neurospora intermedia*, *Penicillium purpurogenum*, *Penicillium canescens*, *Penicillium solitum*, *Penicillium funiculosum*, *Phanerochaete chrysosporium*, *Phlebia radiate*, *Pleurotus eryngii*, *Talaromyces flavus*, *Thielavia terrestris*, *Trametes villosa*, *Trametes versicolor*, *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei*, and *Trichoderma viride*.

The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying the nucleic acid sequence encoding the variant CBH I polypeptide. Culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art. As noted, many references are available for the culture and production of many cells, including cells of bacterial and fungal origin. Cell culture media in general are set forth in Atlas and Parks (eds.), 1993, *The Handbook of Microbiological Media*, CRC Press, Boca Raton, Fla., which is incorporated herein by reference. For recombinant expression in filamentous fungal cells, the cells are cultured in a standard medium containing physiological salts and nutrients, such as described in Pourquie et al., 1988, *Biochemistry and Genetics of Cellulose Degradation*, eds. Aubert, et al., Academic Press, pp. 71-86; and Ilmen et al., 1997, *Appl. Environ. Microbiol.* 63:1298-1306. Culture conditions are also standard, e.g., cultures are incubated at 20 28° C. in shaker cultures or fermenters until desired levels of 25 variant CBH I expression are achieved. Preferred culture conditions for a given filamentous fungus may be found in the scientific literature and/or from the source of the fungi such as the American Type Culture Collection (ATCC). After fungal 30 growth has been established, the cells are exposed to conditions effective to cause or permit the expression of a variant CBH I.

In cases where a variant CBH I coding sequence is under 35 the control of an inducible promoter, the inducing agent, e.g., a sugar, metal salt or antibiotics, is added to the medium at a concentration effective to induce variant CBH I expression.

In one embodiment, the recombinant cell is an *Aspergillus niger*, which is a useful strain for obtaining overexpressed polypeptide. For example *A. niger* var. *awamori* dgr246 is 40 known to produce elevated amounts of secreted cellulases (Goedegebuur et al., 2002, *Curr. Genet.* 41:89-98). Other strains of *Aspergillus niger* var *awamori* such as GCDAP3, GCDAP4 and GAP3-4 are known (Ward et al., 1993, *Appl. Microbiol. Biotechnol.* 39:738-743).

In another embodiment, the recombinant cell is a *Trichoderma reesei*, which is a useful strain for obtaining overexpressed polypeptide. For example, RL-P37, described by Sheir-Neiss et al., 1984, *Appl. Microbiol. Biotechnol.* 20:46-53, is known to secrete elevated amounts of cellulase enzymes. Functional equivalents of RL-P37 include *Trichoderma reesei* strain RUT-C30 (ATCC No. 56765) and strain QM9414 (ATCC No. 26921). It is contemplated that these strains would also be useful in overexpressing variant CBH I polypeptides.

Cells expressing the variant CBH I polypeptides of the disclosure can be grown under batch, fed-batch or continuous 60 fermentations conditions. Classical batch fermentation is a closed system, wherein the compositions of the medium is set at the beginning of the fermentation and is not subject to artificial alternations during the fermentation. A variation of the batch system is a fed-batch fermentation in which the substrate is added in increments as the fermentation

progresses. Fed-batch systems are useful when catabolite repression is likely to inhibit the metabolism of the cells and where it is desirable to have limited amounts of substrate in the medium. Batch and fed-batch fermentations are common and well known in the art. Continuous fermentation is an open system where a defined fermentation medium is added continuously to a bioreactor and an equal amount of conditioned medium is removed simultaneously for processing. Continuous fermentation generally maintains the cultures at a constant high density where cells are primarily in log phase growth. Continuous fermentation systems strive to maintain steady state growth conditions. Methods for modulating nutrients and growth factors for continuous fermentation processes as well as techniques for maximizing the rate of product formation are well known in the art of industrial microbiology.

#### Recombinant Expression in Plants

The disclosure provides transgenic plants and seeds that recombinantly express a variant CBH I polypeptide. The disclosure also provides plant products, e.g., oils, seeds, leaves, extracts and the like, comprising a variant CBH I polypeptide.

The transgenic plant can be dicotyledonous (a dicot) or monocotyledonous (a monocot). The disclosure also provides methods of making and using these transgenic plants and seeds. The transgenic plant or plant cell expressing a variant CBH I can be constructed in accordance with any method known in the art. See, for example, U.S. Pat. No. 6,309,872. *T. reesei* CBH I has been successfully expressed in transgenic tobacco (*Nicotiana tabacum*) and potato (*Solanum tuberosum*). See Hooker et al., 2000, in Glycosyl Hydrolases for Biomass Conversion, ACS Symposium Series, Vol. 769, Chapter 4, pp. 55-90.

In a particular aspect, the present disclosure provides for the expression of CBH I variants in transgenic plants or plant organs and methods for the production thereof. DNA expression constructs are provided for the transformation of plants with a nucleic acid encoding the variant CBH I polypeptide, preferably under the control of regulatory sequences which are capable of directing expression of the variant CBH I polypeptide. These regulatory sequences include sequences capable of directing transcription in plants, either constitutively, or in stage and/or tissue specific manners.

The expression of variant CBH I polypeptides in plants can be achieved by a variety of means. Specifically, for example, technologies are available for transforming a large number of plant species, including dicotyledonous species (e.g., tobacco, potato, tomato, *Petunia*, *Brassica*) and monocot species. Additionally, for example, strategies for the expression of foreign genes in plants are available. Additionally still, regulatory sequences from plant genes have been identified that are serviceable for the construction of chimeric genes that can be functionally expressed in plants and in plant cells (e.g., Klee, 1987, *Am. Rev. of Plant Phys.* 38:467-486; Clark et al., 1990, *Virology* 179(2):640-7; Smith et al., 1990, *Mol. Gen. Genet.* 224(3):477-81).

The introduction of nucleic acids into plants can be achieved using several technologies including transformation with *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes*. Non-limiting examples of plant tissues that can be transformed include protoplasts, microspores or pollen, and explants such as leaves, stems, roots, hypocotyls, and cotyls. Furthermore, DNA encoding a variant CBH I can be introduced directly into protoplasts and plant cells or tissues by microinjection, electroporation, particle bombardment, and direct DNA uptake.

Variant CBH I polypeptides can be produced in plants by a variety of expression systems. For instance, the use of a

constitutive promoter such as the 35S promoter of Cauliflower Mosaic Virus (Guilley et al., 1982, *Cell* 30:763-73) is serviceable for the accumulation of the expressed protein in virtually all organs of the transgenic plant. Alternatively, promoters that are tissue-specific and/or stage-specific can be used (Higgins, 1984, *Annu. Rev. Plant Physiol.* 35:191-221; Shotwell and Larkins, 1989, In: *The Biochemistry of Plants* Vol. 15 (Academic Press, San Diego: Stumpf and Conn, eds.), p. 297), permit expression of variant CBH I polypeptides in a target tissue and/or during a desired stage of development.

#### Compositions Of Variant Cbh I Polypeptides

In general, a variant CBH I polypeptide produced in cell culture is secreted into the medium and may be purified or isolated, e.g., by removing unwanted components from the cell culture medium. However, in some cases, a variant CBH I polypeptide may be produced in a cellular form necessitating recovery from a cell lysate. In such cases the variant CBH I polypeptide is purified from the cells in which it was produced using techniques routinely employed by those of skill in the art. Examples include, but are not limited to, affinity chromatography (Van Tilburgh et al., 1984, *FEBS Lett.* 169 (2):215-218), ion-exchange chromatographic methods (Goyal et al., 1991, *Bioresource Technology*, 36:37-50; Fliess et al., 1983, *Eur. J. Appl. Microbiol. Biotechnol.* 17:314-318; Bhikhambhai et al., 1984, *J. Appl. Biochem.* 6:336-345; Ellouz et al., 1987, *Journal of Chromatography*, 396:307-317), including ion-exchange using materials with high resolution power (Medve et al., 1998, *J. Chromatography A*, 808:153-165), hydrophobic interaction chromatography (Tomaz and Queiroz, 1999, *J. Chromatography A*, 865:123-128), and two-phase partitioning (Brumbauer et al., 1999, *Bioseparation* 7:287-295).

The variant CBH I polypeptides of the disclosure are suitably used in cellulase compositions. Cellulases are known in the art as enzymes that hydrolyze cellulose (beta-1,4-glucan or beta D-glucosidic linkages) resulting in the formation of glucose, cellobiose, cellobiosaccharides, and the like. Cellulase enzymes have been traditionally divided into three major classes: endoglucanases ("EG"), exoglucanases or cellobiohydrolases (EC 3.2.1.91) ("CBH") and beta-glucosidases (EC 3.2.1.21) ("BG") (Knowles et al., 1987, *TIBTECH* 5:255-261; Schulein, 1988, *Methods in Enzymology* 160(25):234-243).

Certain fungi produce complete cellulase systems which include exo-cellobiohydrolases or CBH-type cellulases, endoglucanases or EG-type cellulases and (3-glucosidases or BG-type cellulases (Schulein, 1988, *Methods in Enzymology* 160(25):234-243). Such cellulase compositions are referred to herein as "whole" cellulases. However, sometimes these systems lack CBH-type cellulases and bacterial cellulases also typically include little or no CBH-type cellulases. In addition, it has been shown that the EG components and CBH components synergistically interact to more efficiently degrade cellulose. See, e.g., Wood, 1985, *Biochemical Society Transactions* 13(2):407-410.

The cellulase compositions of the disclosure typically include, in addition to a variant CBH I polypeptide, one or more cellobiohydrolases, endoglucanases and/or β-glucosidases. In their crudest form, cellulase compositions contain the microorganism culture that produced the enzyme components. "Cellulase compositions" also refers to a crude fermentation product of the microorganisms. A crude fermentation is preferably a fermentation broth that has been separated from the microorganism cells and/or cellular debris (e.g., by centrifugation and/or filtration). In some cases, the enzymes in the broth can be optionally diluted, concentrated, partially purified or purified and/or dried. The variant CBH I polypep-

tide can be co-expressed with one or more of the other components of the cellulase composition or it can be expressed separately, optionally purified and combined with a composition comprising one or more of the other cellulase components.

When employed in cellulase compositions, the variant CBH I is generally present in an amount sufficient to allow release of soluble sugars from the biomass. The amount of variant CBH I enzymes added depends upon the type of biomass to be saccharified which can be readily determined by the skilled artisan. In certain embodiments, the weight percent of variant CBH I polypeptide is suitably at least 1, at least 5, at least 10, or at least 20 weight percent of the total polypeptides in a cellulase composition. Exemplary cellulase compositions include a variant CBH I of the disclosure in an amount ranging from about 1 to about 20 weight percent, from about 1 to about 25 weight percent, from about 5 to about 20 weight percent, from about 5 to about 25 weight percent, from about 5 to about 30 weight percent, from about 5 to about 35 weight percent, from about 5 to about 40 weight percent, from about 5 to about 45 weight percent, from about 5 to about 50 weight percent, from about 10 to about 20 weight percent, from about 10 to about 25 weight percent, from about 10 to about 30 weight percent, from about 10 to about 35 weight percent, from about 10 to about 40 weight percent, from about 10 to about 45 weight percent, from about 10 to about 50 weight percent, from about 15 to about 20 weight percent, from about 15 to about 25 weight percent, from about 15 to about 30 weight percent, from about 15 to about 35 weight percent, from about 15 to about 40 weight percent, from about 15 to about 45 weight percent, or from about 15 to about 50 weight percent of the total polypeptides in the composition.

#### Utility of Variant CBH I Polypeptides

It can be appreciated that the variant CBH I polypeptides of the disclosure and compositions comprising the variant CBH I polypeptides find utility in a wide variety applications, for example detergent compositions that exhibit enhanced cleaning ability, function as a softening agent and/or improve the feel of cotton fabrics (e.g., "stone washing" or "biopolishing"), or in cellulase compositions for degrading wood pulp into sugars (e.g., for bio-ethanol production). Other applications include the treatment of mechanical pulp (Pere et al., 1996, Tappi Pulping Conference, pp. 693-696 (Nashville, Tenn., Oct. 27-31, 1996)), for use as a feed additive (see, e.g., WO 91/04673) and in grain wet milling.

#### Saccharification Reactions

Ethanol can be produced via saccharification and fermentation processes from cellulosic biomass such as trees, herbaceous plants, municipal solid waste and agricultural and forestry residues. However, the ratio of individual cellulase enzymes within a naturally occurring cellulase mixture produced by a microbe may not be the most efficient for rapid conversion of cellulose in biomass to glucose. It is known that endoglucanases act to produce new cellulose chain ends which themselves are substrates for the action of cellobiohydrolases and thereby improve the efficiency of hydrolysis of the entire cellulase system. The use of optimized cellobiohydrolase activity may greatly enhance the production of ethanol.

Cellulase compositions comprising one or more of the variant CBH I polypeptides of the disclosure can be used in saccharification reaction to produce simple sugars for fermentation. Accordingly, the present disclosure provides methods for saccharification comprising contacting biomass with a cellulase composition comprising a variant CBH I

polypeptide of the disclosure and, optionally, subjecting the resulting sugars to fermentation by a microorganism.

The term "biomass," as used herein, refers to any composition comprising cellulose (optionally also hemicellulose and/or lignin). As used herein, biomass includes, without limitation, seeds, grains, tubers, plant waste or byproducts of food processing or industrial processing (e.g., stalks), corn (including, e.g., cobs, stover, and the like), grasses (including, e.g., Indian grass, such as *Sorghastrum nutans*; or, switchgrass, e.g., *Panicum* species, such as *Panicum virgatum*), wood (including, e.g., wood chips, processing waste), paper, pulp, and recycled paper (including, e.g., newspaper, printer paper, and the like). Other biomass materials include, without limitation, potatoes, soybean (e.g., rapeseed), barley, rye, oats, wheat, beets, and sugar cane bagasse.

The saccharified biomass (e.g., lignocellulosic material processed by enzymes of the disclosure) can be made into a number of bio-based products, via processes such as, e.g., microbial fermentation and/or chemical synthesis. As used herein, "microbial fermentation" refers to a process of growing and harvesting fermenting microorganisms under suitable conditions. The fermenting microorganism can be any microorganism suitable for use in a desired fermentation process for the production of bio-based products. Suitable fermenting microorganisms include, without limitation, filamentous fungi, yeast, and bacteria. The saccharified biomass can, for example, be made it into a fuel (e.g., a biofuel such as a bioethanol, biobutanol, biomethanol, a biopropanol, a biodiesel, a jet fuel, or the like) via fermentation and/or chemical synthesis. The saccharified biomass can, for example, also be made into a commodity chemical (e.g., ascorbic acid, isoprene, 1,3-propanediol), lipids, amino acids, polypeptides, and enzymes, via fermentation and/or chemical synthesis.

Thus, in certain aspects, the variant CBH I polypeptides of the disclosure find utility in the generation of ethanol from biomass in either separate or simultaneous saccharification and fermentation processes. Separate saccharification and fermentation is a process whereby cellulose present in biomass is saccharified into simple sugars (e.g., glucose) and the simple sugars subsequently fermented by microorganisms (e.g., yeast) into ethanol. Simultaneous saccharification and fermentation is a process whereby cellulose present in biomass is saccharified into simple sugars (e.g., glucose) and, at the same time and in the same reactor, microorganisms (e.g., yeast) ferment the simple sugars into ethanol.

Prior to saccharification, biomass is preferably subject to one or more pretreatment step(s) in order to render cellulose material more accessible or susceptible to enzymes and thus more amenable to hydrolysis by the variant CBH I polypeptides of the disclosure.

In an exemplary embodiment, the pretreatment entails subjecting biomass material to a catalyst comprising a dilute solution of a strong acid and a metal salt in a reactor. The biomass material can, e.g., be a raw material or a dried material. This pretreatment can lower the activation energy, or the temperature, of cellulose hydrolysis, ultimately allowing higher yields of fermentable sugars. See, e.g., U.S. Pat. Nos. 6,660,506; 6,423,145.

Another exemplary pretreatment method entails hydrolyzing biomass by subjecting the biomass material to a first hydrolysis step in an aqueous medium at a temperature and a pressure chosen to effectuate primarily depolymerization of hemicellulose without achieving significant depolymerization of cellulose into glucose. This step yields a slurry in which the liquid aqueous phase contains dissolved monosaccharides resulting from depolymerization of hemicellulose, and a solid phase containing cellulose and lignin. The slurry is then subject to a second hydrolysis step under conditions that allow a major portion of the cellulose to be depolymerized, yielding a liquid aqueous phase containing dissolved/

soluble depolymerization products of cellulose. See, e.g., U.S. Pat. No. 5,536,325.

A further exemplary method involves processing a biomass material by one or more stages of dilute acid hydrolysis using about 0.4% to about 2% of a strong acid; followed by treating the unreacted solid lignocellulosic component of the acid hydrolyzed material with alkaline delignification. See, e.g., U.S. Pat. No. 6,409,841. Another exemplary pretreatment method comprises prehydrolyzing biomass (e.g., lignocellulosic materials) in a prehydrolysis reactor; adding an acidic liquid to the solid lignocellulosic material to make a mixture; heating the mixture to reaction temperature; maintaining reaction temperature for a period of time sufficient to fractionate the lignocellulosic material into a solubilized portion containing at least about 20% of the lignin from the lignocellulosic material, and a solid fraction containing cellulose; separating the solubilized portion from the solid fraction, and removing the solubilized portion while at or near reaction temperature; and recovering the solubilized portion. The cellulose in the solid fraction is rendered more amenable to enzymatic digestion. See, e.g., U.S. Pat. No. 5,705,369. Further pretreatment methods can involve the use of hydrogen peroxide H<sub>2</sub>O<sub>2</sub>. See Gould, 1984, Biotech, and Bioengr. 26:46-52.

Pretreatment can also comprise contacting a biomass material with stoichiometric amounts of sodium hydroxide and ammonium hydroxide at a very low concentration. See Teixeira et al., 1999, Appl. Biochem. and Biotech. 77-79:19-34. Pretreatment can also comprise contacting a lignocellulose with a chemical (e.g., a base, such as sodium carbonate or potassium hydroxide) at a pH of about 9 to about 14 at moderate temperature, pressure, and pH. See PCT Publication WO2004/081185.

Ammonia pretreatment can also be used. Such a pretreatment method comprises subjecting a biomass material to low ammonia concentration under conditions of high solids. See, e.g., U.S. Patent Publication No. 20070031918 and PCT publication WO 06/110901.

#### Detergent Compositions Comprising Variant CBH I Proteins

The present disclosure also provides detergent compositions comprising a variant CBH I polypeptide of the disclosure. The detergent compositions may employ besides the variant CBH I polypeptide one or more of a surfactant, including anionic, non-ionic and ampholytic surfactants; a hydrolase; a bleaching agents; a bluing agent; a caking inhibitors; a solubilizer; and a cationic surfactant. All of these components are known in the detergent art.

The variant CBH I polypeptide is preferably provided as part of cellulase composition. The cellulase composition can be employed from about 0.00005 weight percent to about 5 weight percent or from about 0.0002 weight percent to about 2 weight percent of the total detergent composition. The cellulase composition can be in the form of a liquid diluent, granule, emulsion, gel, paste, and the like. Such forms are known to the skilled artisan. When a solid detergent composition is employed, the cellulase composition is preferably formulated as granules.

#### EXAMPLES

##### Materials and Methods

##### Preparation Of CBH I Polypeptides For Biochemical Characterization

Protein expression was carried out in an *Aspergillus niger* host strain that had been transformed using PEG-mediated transformation with expression constructs for CBHI that

included the hygromycin resistance gene as a selectable marker, in which the full length CBH I sequences (signal sequence, catalytic domain, linker and cellulose binding domain) were under the control of the glyceraldehyde-3-phosphate dehydrogenase (gpd) promoter. Transformants were selected on the regeneration medium based on resistance to hygromycin. The selected transformants were cultured in *Aspergillus* salts medium, pH 6.2 supplemented with the antibiotics penicillin, streptomycin, and hygromycin, and 80 g/L glycerol, 20 g/L soytone, 10 mM uridine, 20 g/L MES in baffled shake flasks at 30° C., 170 rpm. After five days of incubation, the total secreted protein supernatant was recovered, and then subjected to hollow fiber filtration to concentrate and exchange the sample into acetate buffer (50 mM NaAc, pH 5). CBH I protein represented over 90% of the total protein in these samples. Protein purity was analyzed by SDS-PAGE. Protein concentration was determined by gel densitometry and/or HPLC analysis. All CBH I protein concentrations were normalized before assay and concentrated to 1-2.5 mg/ml.

##### CBH I Activity Assays

**4-Methylumbelliferyl Lactoside (4-MUL) Assay:** This assay measures the activity of CBH I on the fluorogenic substrate 4-MUL (also known as MUL). Assays were run in a costar 96-well black bottom plate, where reactions were initiated by the addition of 4-MUL to enzyme in buffer (2 mM 4-MUL in 200 mM MES pH 6). Enzymatic rates were monitored by fluorescent readouts over five minutes on a SPEC-TRAMAX™ plate reader (ex/em 365/450 nm). Data in the linear range was used to calculate initial rates (V<sub>o</sub>).

**Phosphoric Acid Swollen Cellulose (PASC) Assay:** This assay measures the activity of CBH I using PASC as the substrate. During the assay, the concentration of PASC is monitored by a fluorescent signal derived from calcofluor binding to PASC (ex/em 365/440 nm). The assay is initiated by mixing enzyme (15 µl) and reaction buffer (85 µl of 0.2% PASC, 200 mM MES, pH 6), and then incubating at 35° C. while shaking at 225 RPM. After 2 hours, one reaction volume of calcofluor stop solution (100 µg/ml in 500 mM glycine pH 10) is added and fluorescence read-outs obtained (ex/em 365/440 nm).

**Bagasse Assay:** This assay measures the activity of CBH I on bagasse, a lignocellulosic substrate. Reactions were run in 10 ml vials with 5% dilute acid pretreated bagasse (250 mg solids per 5 ml reaction). Each reaction contained 4 mg CBH I enzyme/g solids, 200 mM MES pH 6, kanamycin, and chloramphenicol. Reactions were incubated at 35° C. in hybridization incubators (Robbins Scientific), rotating at 20 RPM. Time points were taken by transferring a sample of homogenous slurry (150 µl) into a 96-well deep well plate and quenching the reaction with stop buffer (450 µl of 500 mM sodium carbonate, pH 10). Time point measurements were taken every 24 hours for 72 hours.

**Cellobiose Tolerance Assays (or Cellobiose Inhibition Assays):** Tolerance to cellobiose (or inhibition caused by cellobiose) was tested in two ways in the CBH I assays. A direct-dose tolerance method can be applied to all of the CBH I assays (i.e., 4-MUL, PASC, and/or bagasse assays), and entails the exogenous addition of a known amount of cellobiose into assay mixtures. A different indirect method entails the addition of an excess amount of β-glucosidase (BG) to PASC and bagasse assays (typically, 1 mg β-glucosidase/g solids loaded). BG will enzymatically hydrolyze the cellobiose generated during these assays; therefore, CBH I activity in the presence of BG can be taken as a measure of activity in the absence of cellobiose. Furthermore, when activity in the presence and absence of BG are similar, this indicates toler-

ance to cellobiose. Notably, in cases where BG activity is undesired, but may be present in crude CBH I enzyme preparations, the BG inhibitor gluconolactone can be added into CBH I assays to prevent cellobiose breakdown.

#### Library Screening Assays

The wild type CBH I polypeptide BD29555 was mutagenized to identify variants with improved product tolerance. A small (60-member) library of BD29555 variants was designed to identify variant CBH I polypeptides with reduced product inhibition. This product-release-site library was designed based on residues directly interacting with the cellobiose product in an attempt to identify variants with weakened interactions with cellobiose from which the product would be released more readily than the wild type enzyme. The 60-member evolution library contained wild-type residues and mutations at positions R273, W405, and R422 of BD29555 (SEQ ID NO:1), and included the following substitutions: R273 (WT), R273Q, R273K, R273A, W405 (WT), W405Q, W405H, R422 (WT), R422Q, R422K, R422L, and R422E (4 variants at position 273×3 variants at position 405×5 variants at position 422 equals 60 variants in total). All members of the library were screened using the 4-MUL assay in the presence and absence of 250 g/L cellobiose and using gluconolactone to inhibit any BG activity. The R273A, R273Q, and R273K/R422K variants showed enhanced product tolerance. The R273K/R422K variant showed greatest activity among the variants and cellobiose tolerance at 250 mg/L. Due to low expression, the R273K variant was not tested for product inhibition.

#### Characterization of Product Tolerant Variants of BD29555

The R273K/R422K substitutions were characterized in both a wild type BD29555 background and also in combination with the substitutions Y274Q, D281K, Y410H, P411G, which were identified in a screen of an expanded product release site evolution library.

The wild type, the R273K/R422K variant and the R273K/Y274Q/D281K/Y410H/P411G/R422K variants were tested for activity on 4-MUL in the presence and absence of 250 mg/L cellobiose, and the R273K/R422K variant was also tested in the bagasse assay in the presence and absence of BG. The results are summarized in Table 5.

The results from these activity assays were converted into the percentage of activity remaining with and without cellobiose present, where values close to 100% indicated cellobiose tolerance. The percent of activity remaining in the MUL assay in the presence cellobiose versus in the absence of cellobiose shows that the R273K/R422K variant was the most tolerant, followed by the R273K/Y274Q/D281K/Y410H/P411G/R422K variant, and then wild-type, at 95%, 78%, and 25% activity, respectively.

Cellobiose dose response curves of the wild-type and R273K/R422K variant of BD29555 were obtained during the 4-MUL assay. Enzyme rates ( $V_o$ ) were measured in the presence of different concentrations of cellobiose (200 mM MES pH 6, 25° C.). Rates were measured in quadruplicate. The results are shown in FIG. 1A-1B. FIG. 1A shows that wild type BD29555 is inhibited by cellobiose, with a half maximal inhibitory concentration ( $IC_{50}$  value) of 60 mg/L. FIG. 1B shows that the R273K/R422K variant is tolerant to cellobiose up to 250 mg/L.

The bagasse assay results shown in Table 5, which lists the percentage of activity remaining in the absence vs. presence of BG, also demonstrate that the percentage activity of the

wild type BD29555 is lower than the percentage activity of the R273K/R422K variant, indicating that the R273K/R422K variant is less sensitive to the presence of cellobiose than the wild type. FIG. 2A-2B shows bar graph data for the bagasse assay of BD29555 vs. the R273K/R422K variant. In FIG. 2A, bars represent relative activity, which has been normalized to wild type activity in the absence of cellobiose (WT+BG=uninhibited activity=1). In FIG. 2B, bars indicate tolerance to cellobiose, as represented by the ratio of activity in the presence of cellobiose (-BG) to that of activity in the absence of cellobiose (+BG); ratios close to 1 indicate greater tolerance to cellobiose. These data again demonstrate that the R273K/R422K variant of BD29555 is more tolerant to cellobiose than the wild type BD29555.

The wild type and R273K/R422K variant were also characterized in the PASC assay. Results are shown in FIG. 3. The activities of both wild type BD29555 (SEQ ID NO:1) and wild type *T. reesei* CBH I (SEQ ID NO:2) were inhibited by cellobiose concentrations starting around 1 g/L (with  $IC_{50}$  values of 2.2 and 3 g/L, respectively), whereas the R273K/R422K variant showed little inhibition in the presence of 10 g/L cellobiose.

#### Characterization of Product Tolerant Variants of *T. reesei* CBH I

Cellobiose product tolerant substitutions were introduced into *T. reesei* CBH I (SEQ ID NO:2). A panel of variants with single and double alanine and lysine substitutions at R268 and R411 were expressed and analyzed. The variants were tested for activity on 4-MUL in the presence and absence of 250 mg/L cellobiose and also in the bagasse assay in the absence and presence of BG. The results from these assays were converted into the percentage activity remaining in the presence and absence of cellobiose and BG, respectively. Values are summarized in Table 6.

The 4-MUL assay results shown in Table 6 demonstrate that the activity of the wild type *T. reesei* CBH I was reduced to 23% in the presence of cellobiose, whereas the double mutants at R268 and R411 retained more than 90% of their activity under the same conditions.

The bagasse assay results shown in Table 6 demonstrate that the activity of the wild type *T. reesei* CBH I is more significantly impacted by the presence of BG than is the activity of the single or double substitution variants, indicating that the variants are less sensitive to the accumulation of cellobiose than the wild type. FIGS. 4 and 5 show bar graph data for the bagasse assay of wild type *T. reesei* CBH I vs. the variants. In FIG. 4, bars represent relative activity, normalized to wild type activity in the absence of cellobiose (WT+BG=1). In FIG. 5, bars represent tolerance to cellobiose, as represented by the ratio of activity in the presence of accumulating cellobiose (-BG) to that of activity in the absence of cellobiose (+BG); ratios close to 1 indicate greater tolerance to cellobiose.

#### Specific Embodiments and Incorporation by Reference

All publications, patents, patent applications and other documents cited in this application are hereby incorporated by reference in their entireties for all purposes to the same extent as if each individual publication, patent, patent application or other document were individually indicated to be incorporated by reference for all purposes.

While various specific embodiments have been illustrated and described, it will be appreciated, that various changes can be made without departing from the spirit and scope of the invention(s).

TABLE 1

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 1	BD29555*	Unknown	MSALINFNNY KSAIILIGSLI ATAGAQOQIT YTAEATHPSLS WSTCKSGGSC TTNSGAITLD ANWRWYHGVN TSTNCYTGTNT WNTAICDIDA SCZQDCALDG ADYSCTGYIT TSGNSLRLNF VTGSNVGERT YLMADNTHYQ IFDLINQEFT FTVIWSHLPIC GINGAL-YFVT MDAGEVSKY PNNKRAQAQYVG VGYCDSQCR DLKFIAGQN VEGMTPSSNN ANTICLGNHGA CCZELD IWEA NSISEALTHPH PCDTFGLSVC TTDACTGNTYS SDRYAGTCDP DGCDNPYRL GVTIPIYGSGK TVIQLVTDGT YVYQNGVYIP QPSKLSGV SPTAGTGPAA GNVINSIDRCF AEISTEGETA SSFKHGHHIAK MGAGMEAGMV LVMSLWDDSY PMATGSPGA ARGSPCTTSQG DPKTVESEOG SSYTFSDIR VGPENSTFSQ GSSTGSSSTT TASGTTTKA SSTSTSSTT GTGVIAAHMGQ CGGGCWTGFT TCZSQTCTV VNPYYSQCL
SEQ ID NO : 2	340514556	Trichoderma reesei	MYRKLAIVSA FLATARAQSA CILLOSETHPP LTWOKCSSGG TCTQGTGSVY IDANMRWTHA TNSSSTNCYDG NTWSSTLCPD NETAKRNCIL DGAAYASTYG VTSGNSLSSI GFTVOSAQKN VGRLYIMAS DITYQETTL GNEFESFDVDY SOLEFGGLNGA LYFVSMADG GVSXKPTNTA GAKYGTGYCD SQCPDLKPI NGOANYEGW PSSNANTGII GHGSOCCSSEM DIWEANSUZE ALTPHP-TTV GOEIEGEGC GGTSNDNYG GTCDPDGWDW PNVRGNTSF YGPSSFTILD TYKVKLTIVTO PETSGAIIYR VYQNGVYTFQ PNAELGSTYSG NEFLNDYCTA EEAERGSSF SDKGCLTQK KATSGGMVLV MSLNDDYAN MLLWLDSTYP NETASLPGAV RESCSTESGV PAQEVQSPN AKYTFENIKP GPJGSTGNPNS GGNPPGNNP GTTTRPRAT TTGSSSPGFTQ SHYGQCQGG YSGPTVCAASG TTCCQVLYNPYY SQCL
SEQ ID NO : 3	51243029	Penicillium occitanis	MSALINFNNY KSAIILIGSLI ATAGAQOQIT YTAEATHPSLS WSTCKSGGSC TTNSGAITLD ANWRWYHGVN TSTNCYTGTNT WNSAIICDIDA SCZQDCALDG ADYSCTGYIT TSGNSLRLNF VTGSNVGSR YLMADNTHYQ IFDLINQEFT FTVIWSHLPIC GINGAL-YFVT MDAGEVSKY PNNKRAQAQYVG VGYCDSQCR DLKFIAGQN VEGMTPSSNN ANTICLGNHGA CCZELD IWEA NSISEALTHPH PCDTFGLSVC TTDACTGNTYS SDRYAGTCDP DGCDNPYRL GVTIPIYGSGK TVIQLVTDGT YVYQNGVYIP QPSKLSGV SPTAGTGPAA GNVINSIDRCF AEISTEGETA SSFKHGHHIAK MGAGMEAGMV LVMSLWDDSY PMATGSPGA ARTGATATTSG DPKTVESEOG SSYTFSDIR VGPENSTFSQ GSSTGSSSTT TASRTTTSA SSTSTSSTT GTGVIAHMGQ CGGGCWTGFT TCZSQTCTV VNPYYSQCL
SEQ ID NO : 4	7cel (PDB) &	Trichoderma reesei	ESACTLQSET HPLITWQKCS SGCTCTQGT SVIDANMRW THATNSSTNC YDNTWWSNL CPDNETAKN CCLDDBAYAS TYGTTSNSN LSDFDVTOA QKNGARYL MASDTYOEF TLIGNEFFED VDVSQLPQGL NGALIVFVSMID AGFVSKPFT NTAGAKYR YCDSQCRDL KFVNQCANM TGJGHGSSC SEMIDWOANS ISEALTHPHC TTNGOEICEG DCGCGNTYSDN RYGETCDDPG CIDNPYRAGN TSFYGPSSP TLDTFLKLTV VTOBETSGAI NRIVYVQNGVYF EQPNALGS YSGNELNDYY CTABAEFGG SSFSDKGSLT QFKKATSGGM VLYNWLDDY YAMWLWLDST YPTNETSSTP GAVRSSCSTS SGVPAQVESQ SPNAKVTFSN IKFGP1GST NPSG
SEQ ID NO : 5	67516425	Aspergillus nidulans FGSC A4	MASSPOLYKA LLFFSSLLSA VQAKVGTQO AEVHPGLTWO VNGEVITDAN WEWLHTVNGY TNCYTGNEND TSTICTSNEWC ARCAVAGAN YASTYGTTS GSSURLNFTV QSOQKNITSR VYLMDDDETY TMFYLIANKEP TFDVDSLELP CGLINGAYIVF SMADGKSR YATNEAGAKY GTGICDSQCP RDLKFLINGVA NVEGFESSDT NPNGGYNHNG SCCAEMDIWE ANSISTAFTP HPCTDPGOTL CTEDSGGTY SNDRYGTCD PDGDENSYR QGNTTFYGP G LVTDTNSPAT VVTQFLTDDN DTGTLUSEIK RFVQNGVVI PNESESTYPA PGNSLTTEPC ESOKELEFDV DYESAHHGMA GMGAALFGM VLVLSLWDN YSNMLWLDN YPTDADPQP GIARGRCPTD SGVSEVEAQ YPAVYVSN IRFGPIGSTP GNGGSSGPTT TVTISTATST TSATSTATG QAQHMQQCGG NGHTGPTVCA SPWACTVNS WYSQCL
SEQ ID NO : 6	46107376	Gibberella zeae PH-1	MYRALATASA LIAAVRAQV CSITOESKPS LNWSKCTSSG CSNVKGSVTI DANWRWTHQV SGSTNCYTG KWDITVCTSG KVCELEKCCUD GAYASTYGI TSISGDQSLIS FVTKGPSYTN IGERTYLMED ENTYQMQLL GNEFESFDVDY SMICGGLNGA LYFVSMADG GKAHYQKNA GAKYJGQKNA ACPRDVKEI NGQANSQGW PSDSDVNGGI GNLTGCTCPM DIWEANSIST AYTHPCTKL TQHSCTGDSC GGTYSNDRYQ GTCDADGDF

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 7	70992391	<i>Aspergillus fumigatus</i> Af293	NSYRCNKT YGPSSGFNYD TTKVTVVTO FHKSGNRSL EITRLYVONG KVIANSESKI AGVPGNSLTA DFTCFKRVF NDPPDFTRK AVGMSDLW APMVLMWS HDHNSNMWL DSTPTSDST LGSQRGCSQT GSPVADLEK NVPISKVAFS NIKFGPIGST YKSDGTTPN PTNPSEPSNT ANPPIGTVDO GPTAKSGFT CKKNDFTSQ CO
SEQ ID NO : 8	121699984	<i>Aspergillus clavatus</i> NRRL 1	MLASTPSYR YKTLILAL LGGGQAQOVG TSOAEVHPSM TWQSCTAGS CTTNGKRVVI DANWRWTHKV GDYTMTCYTG TWDTTICPND ATCANSQALE GANVESTYGV TASGNSLRLN FVTSQQRNI GSRLYMMKDD STYENFKLLN QEFFFDVDS MLLCGLNLAL YFVAMDADEGG MISKPYTKAG AKYGTGCVDS QCPRLKFPIN GOANTEGMOP SSNDANANGT NHGCCAEMD IWEANSTSTA FTPHPCDTPG QVMTGDAEG CTYSSDREYGG TCDPGCDEF SFRGANKTYI GPWTVDIKS KFTVVTQFIT DDGHTSGTLK EIRKFYVONG KVIPNSBS-TW TGVSNSITI EYCAQAKSUF ODQNVFERKG GIEGMGAALA EGMLVLSLN DDHSANMMLW DSNYPITAS S TTPGARGTC DISCGVPAVD EAHNHPDAVV YSNIKVGPIG STENSGGSNP NGGTTTTTTT QPTTTTTAG NGGTCVAQH YGQCGGIGT GPTCA-ASP-T CQLNDYYSQ CL
SEQ ID NO : 9	1906845	<i>Claviceps purpurea</i>	MLPSTISYR YKNALFFAAL FGGVQAQKVG TSKAEVHPSM AWOTCAADGT CTTPKNGKVVI DANWRWTHDV KGTYTCYTG TWNELCPDN ESGAENCAL GADYAAATGYA TTSGNALSK FVTSQQRNI GSRLYMMKDD NTYEFKLLN QEFFFDVDS NILFGCLNGAL YFVSMNDADGG LSRTGNEAG AKYGTGCVDS QCPRLKFPIN GLANTEGMWP SSNDANANGT GHSSCCAEMD IWEANSTSTA YFVTSQQRNI GSRLYMMKDD QMCPYDSCG CTYSSDREYGG TCDPGCDEN SYRQGNKSPY CGMTVDTKK KMVTVQFELT YDPAFTGTLK EIRKFYVONG KVIANSESTW PNLLGNSLTN DFCQAQKTVF GMDTFSRKH GMEGMGRALA EGMLVLSLN DDHSANMMLW DSNYPITAS S TTPGARGSC DISSGDPKD1 EAHNPDASVY YSNIKVGPIG STENSGGSNP GGSTTTTKPA TSTTTKAT TATTNTTGT GTGTAQPWQ CGGIGSCEPT QCAAPYCTK QNDYTSQCL
SEQ ID NO : 10	19p1 (PDB)	& <i>Phanerochaete chrysosporium</i>	MHPSLQTTI SALFTTAHQ QACSSKPETH PPLSWSRCR SGCRSVQAV TVDANWLTAT VDGSONCYTG NRWDSCISS EKTOSESCII DGDYADGTYC VTTTEDALSL KEYQOGPKS IVGSRPLYLMK DESRYEMETL LGNETTFDVS DVKLGCGCLN ALYFVMSNDE GGMKREPMLN AGAKFGTGYC DSCOPRDYKF INGMANSKDW IPSKSDANAG TGSIGACCRE MDIWEANNDI GSFTHPKCKN SAYHSGTCDG CGGTTSYKRY SGDCCDPDGCD FNSYVLGNTT FYGEGPKFTI DTRKLSVVT QFLKGDGSL REIKEFVYON GKYLPSYSR VRGVPGNSIT T QGFCNAQKRM FGAHESENKA GGMKGMSAAV SKPMVLMVMSL WDDHSNMWL LDSTYPTNSR QRGSKRSQP ASSGPRTDVE SSADSTVVF SNIKFGPIGS TEFRGK
SEQ ID NO : 11	119468034	<i>Neosartorya fischeri</i> NRRL 181	ESACTLOSET HPPLTWOKCS SGCTCTOOTG SVVIDANMRW THATNSSTNC YDENTWSSTL CPDNETCAKN CCLDZAYAS TYGTTGNS LSIDFVTOSA QKVNKGARLYL MASDTYQEF TLLGNEFSTD VDVSQLPQGL NGALFEVSM ADGGVSKYPT NTGAKYCTG YODSOCPRDL KFINQOANE GWEFSSNDAN TGJGGHESCC SEMDIWQANS ISELALPHC TTGQEIEG DCGGTYSNDN RGFTCDPDG TLDITKKLTV VTOEBSGAI NEYVQONCVT FQOPNEELGS YSGNELNDYY CTABEAERFG SSFSDKGQLT QPKKATSGGM VLWMSLWDYY YAMLWLDST YPTNETSSTP GAVRGSCTS IKFGTGSTQ NPSC
SEQ ID NO : 12	7804883	<i>Leptosphaeria maculans</i>	MHQALLFSA LAYVANAOQV GPKPTEHPP LTWOKCTAAG SCQOQSSTV IDANWRWLSHs TKDTTNCYTG NTWANELCPD NESAQNCANV DGAHYAGTYG VITSGSELKL SFVTCANVGVS RLYMQDDET YOHNLLNNE FTEFDVDSNL PCGNGALYF VANDADGEMS KYPNSNAGAK YGTGTCDSOC PRDLKFTGM ANVEGMIPS S NDKNAGVGGH GSCCEPEMDIW EANSISTAVT PHPCDDVQSQT MCGFAAGGT YSATRYAGTC DEDGCDNPFP RMGNFSFYGP GKTIVDTKSEM TVTQFLTAD GDTGTAUSEI KRLVYQONGKV TANVSINYAD VSGNSLISSDP CTAQKARGD EDIIFAKGEL SNGKALUSEM VLIMSIVDDH HSMMWLDST YPTADPSKP GVARGTCBHG AGDPKEVESQ HPDSAVTFEN IKRGPISSTY KA
			MYRSILJFATS LLSLAKGOLV GMLYCKGSCT AKNGKVVIDA NWRMLHVKGG YTNCYTGNEW NATACPDKS CATNAGA DYLRLRHYCE RQLGTEVHH QCLYSTNIGS RTYLMQDST YQLFKFTGQ EFTEVDLNSN LPCGNGALY FVSMNDADGEL KTKPTNITAGA KYGTGTYCDAQ CPRDLKFTING EGNVVEGWQPS KNDQNA-GYGG

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 13	85108032	<i>Neurospora crassa</i> N150 (OR7A)	<p>HGSCCAEMDI WEANISVSTAV TPISCSLEQ SRCDGDCGG TYSADRYAGV CDPDGCDENS YRMGVKDFYIG      KGKTYDTTSKK FTVTQFLGS GDAEHTIPQP VONGKTPQF DSTITPGVY SITTEFFCDAQ KRAFGDFTF      KDKGWNANMP STCGMVLYM SLWDDHTSNN LWLDDSTYPTD KNPDTDAGSG RGECAITSQV PADVEQSHPD      ASVIVSNIKF GPINTTGF</p> <p>MLAKPAALAA LVASANAAQAV CSLTAETHPS LNWKCTSSG CTVAGASITV DAMWRWHIT SGSTNCYSGN      EMDTSLCSTN TDCTTKCCVD GARYSSSTYGI QTSGENSLSLQ FVTYGSYSTN IGERTYLIMNG ADAYOGHELL      GNEFTFDVY SGTCGGLNSA LYTFMSMDLG GRAKYTNNKA GAKYGTGYCD ACPRDLYKI NGJIANVBSWT      PSTNINANGI GDHETCCSEM DWEANKVST AFTPHPCPTI EOHMCEGDSC GGTYSSDDEYG GNCDADGCF      NSYRQNTTF YGBGKTVIDUS SKETVUTQFI RDAGSLIAE KRFVYQNGKV VSGNSNVDG VSGNSNVDG VSGNSNVDG      CNAQTAARGD IDDENPKKGSL KONGKALAKP MVLVMSIWDD HAANNMLWLD TYPEVEGGGA YRGECPTISG      VPAEYEAAP NSKTIFSNIK RGIGSTFSG GSISGTPPSNP SSSVPVST AKPSSTAS NPSGTGAAW      AQCGGIGFSGS PTTCOSPYC QKINDYFSQC V</p>
SEQ ID NO : 14	169859458	<i>Coprinopsis cinerea</i> Okayama	<p>MFKKYALTTAL CLEVAQ-AQ VGFEVAENHP RLPIWQRCTR N GGCQTV/SNGQ VVLDANWML HYTDGYTNCY      TGNSPNSTVC SDPHTCAQFC ALIGANTFOOT YGTTINGDAL T KIPULTRSQO TNVARYATLM ENENRYONEN      LLNKEFTEDV DVSKVPCGIN GALYFIQMDA DCGMSKOPNN RAGAKYGTGY CDSSCPRDIK FIDGVANSAD      WTPSETDNA GRGYGICCA EMDIWEANSI SNAYTHPCR TQNDGGYQRC EGDRVANOPT EGICCDPDGCD      YNPFERMGNKD FYGGKTVDTI WRKHMVYTOQ ITHDNNDTGT LDVIRRLYVQ DGRVIANOPT NPPGLMZAHD      SITEOPCTDQ KNLFGDYSSF ARIGGLAHMG RSLAKGHVLA LISWINDHGAH MLWLDNSNPT DADPNKPGIA      RGTCPITGGT PREEQNHPD AQVIESNIKP GDIGSTFSGY</p> <p>MYSAVILATE SPILLGAGQ VGTSTAETHD ALTVOKCAG CTCTDESDSI VLDANWMLH STSGSTNCY      GNTWDITLCP DAATCTTNCIA LDGADYECTY GTTSCBDSLK LSFTUTGSNVG SRATYLMDSST TYKEFALLGN      EFTFTDVSK LPGCLNGALY FVMDADGGM SKYPTPKAGA KYGETGYCDAQ CPQDMKELYNG TANVEGWVPD      SNSANSGTMN IGSCCSEFPD WEANSMSCAL TPHVCTVDSO TAECTGDCAS NTGVCDDGICC DFNPNYRUGNT      TPGCMTID TSKESEVUTQ FTTDGTETG TLTEKREY SD15GVSGNS ITDDPCZAK      TAFTGTDYFT QNGGNAEKG KWDAGMVYLV SIMDDTNVNM LWLDDYPTT KDASTPGVSR GSCATDGV      ATVEALGSA VVTESSIKYG PICSTENAPA DSSSYSSASS SAPTASSSS SASIAPVSV VAISSSSA      AISSAPVVS SSAOAISSAA PVVSSVSSA APVATESTS KCSKVSSTLK TSYAAPATSA TAAVATATS      AASSPSGVPL YGMGGKTC SECTCVQVOND YSSQVASS</p> <p>MTWQRTGTG GSSCTINVNGE IVTDANWELI HATGGYTNCF DGNEWKTAC PSNAACTKNC ALEGSDYRG      YGITSGNSL TLKETIKGQV SNTVGSRSTYL MKDTNMYEMF NLIGNEFTFD VDLSOLPCGL NCALYFTSMP      EKGQGTPGAK YGIGKLQS CARDLKEVGGT BANADQGQAS TSDPNAGVKG KCACCAEMDV      WEANSMSTAL TPHECOPEGY AVCEESNGG TYSLDEYAGT CDAANGDFENP YRGZNKDFYKG KGTKTVDTSK      MTVVFQFLGT GSDLTELKEF YVQDGKVLSN PEPTIPMTG NISITOKWCDT QKEVYKEEVY PPNQWGGMAS      MGKGAQGMV LVMSLWDDH SNMLWLDSTY PTDRDBESPQ AREGCAITS      KFGPFGSTFQ QPA</p>
SEQ ID NO : 15	154292161	<i>Botryotinia fuckeliana</i> B05-10	<p>MQIKSYIQL AAALPILSSV AAOQAGITTA ENHPRMTWKR CSGP-GNCQTV QGEVVIDANW RMILHNNGCNC      YEGNWTISOC SSATDCAQRC ALIGANTOST YGASTSGDSL TLKFVTKHEY GTNIGSREFL MANONKTYOMP      TLMMNEFAFD VDLSKVECGI NSTLIFYVAME EDGGMASYPS NRAGAKYGTG YCDAQCARDL KEIGGRANIE      GMRPSTNDP AGYGPMSGAC AEDWNEENA YAYAFPHAC GSKNRYHICE TNNGGTYSD DRFAGYCDAN      GCDYNPYRMG NKDYGKGTI VDTNTRKFTVV SRFERURLISQ FVQDGRKIE VPPPTWPGLP NSADITPELC      DAQFRVEDDR NREFETGGFD ALINEALTIPIV VLMWSIWDDH HSNNMLWLDSS YPPEKAUGLP GDRGPCPITS      GVPAVEAQV PNAQOVWSNI RFQFPIGSTVN V</p>
SEQ ID NO : 16	169615761 #	<i>haeo-sphaeria nodorum</i> SN15	<p>MRTAKFATLA ALVAAAQQ ACCLTTEHHP SLSWKKCTAG GQCTQVQASI TLDSENWRTWQ QVSGSTNCY      GNKWTDTISCT DAKSCAQNC C VDGADYISTY GITTINGDSLS LKFKVTKGQYS TNVGSRTVLM DGEDKYQTTE</p>
SEQ ID NO : 17	4883502	<i>Humicola grisea</i>	<p>MRTAKFATLA ALVAAAQQ ACCLTTEHHP SLSWKKCTAG GQCTQVQASI TLDSENWRTWQ QVSGSTNCY      GNKWTDTISCT DAKSCAQNC C VDGADYISTY GITTINGDSLS LKFKVTKGQYS TNVGSRTVLM DGEDKYQTTE</p>
SEQ ID NO : 18	9506886	<i>Humicola grisea</i>	<p>MRTAKFATLA ALVAAAQQ ACCLTTEHHP SLSWKKCTAG GQCTQVQASI TLDSENWRTWQ QVSGSTNCY      GNKWTDTISCT DAKSCAQNC C VDGADYISTY GITTINGDSLS LKFKVTKGQYS TNVGSRTVLM DGEDKYQTTE</p>

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 19	124491660	<i>Chaetomium thermophilum</i>	LLGNNRFTEDV DVSNITGGLN GAIYFVSMDA DCGLRSYPGN KAGAKYGTGX CDACOPRDIK FINGEANIEG WTGSNDPNA GAGRYGTCGCS EMPIWEANM ATAFTHPHCT IIGQERPHCT SCGGTYNSER YAGVCDPDGC DFNSTRQGNK TFYKGMTVD TTJKITVYTO FLKDANGDLG EIKREYVQDG KLPNSESTI PVEGENSITQ DWCDQKVFA GDIDENRKG GMQMGKRALA GMVLWMSIW DDHASNMLWL DSTFPVDAAG KPGAERGACP TTSGPAAVE AEANSNTVVF SNIRFGPIGS TVAGIPGAGN GNINNGNPPP PTITTSAPA TTITASGPK AGRWQCGGI GFTGPTQCEB PYTCKLNDW YSQL
SEQ ID NO : 20	58045187	<i>Chaetomium thermophilum</i>	MQIKQYLQL AAALPLVNA AAGRAGTQQT ETIPLRSLWR CSGNGCQTV NAEILVIDANW RFLHDSNYQN CYDGNRWTSA CSSTADCAQK CYLEGANYGS TYGVSTSGDA LTLKEVTKHE YGTNIGSERYV LMNGSDKYOM FTLMNNEFAF DVDSLKVCEG LNSALYEVAM BEDGGMKSYGN SNKAGAKYGT CGDQACARD LKFVGKANI EGMRSTNDNA NAGGPyGAC CABDYNESN AYAFAPTPHG CLNNNNYHVC TSNGGTYSE DRFGGLCDAN GCDTNPYRMG NKDYGKGTI VDTSRKTFIV TRFEENKLTO FFIQGKRID IPPTWPGLP NSAITPLC TNLSKVEDDR DRYETGGER TINBALRIPM VLVMSWLDHV YANMILWLDHV YPPEKAGQPG AERGPCTAPS GVPAAVEAQF PNAQYIWSEN RFGCPGTYQ V
SEQ ID NO : 21	169661100 #	<i>Phaeosphaeria nodorum</i> SN15	MMYKRFALAA ALVYGAAGQ ACCLTETHP RLTMKRCTSG GNCSTVNGAV TIDANWRTH TVSGSTNCTY GNEDWTSIC S DGKSCAQTC VDGDADSYT GITTGSDSLN LKFVTKHOGH TNVGSRVLYM ENDTKYQMF LLGNNRFTEDV DVSNITGGLN DCGMSKYSGN KAGAKYGTGX CDACOPRDIK FINGEANIEG WTPSNDANA GREGYGCSCS EMPIWDANM ATAFTHPHCT IIGQERPHCT SCGGTYSSER YAGVCDPDGC DFNAVRQGDK TFYKGMTVD TTJKITVYTO FHKNSACVLV EIKREYVQDG KITANAESKI PGNPGNTITQ EWCDQKVFA GDIDENRKG GMQMGKRALA GMVLWMSIW DDHYANMLWL DSTPDKAG TPGAERGACP TTSGPAAEIE AQVENSNTVIF SNIRFGPIGS TVPGDGSTP SNPTATVAPP TSTTTSVRS TTQISTPTSQ PGCGTQKNG QCGGIGYTCG TICVAGTCTT ELNPWHSQCL
SEQ ID NO : 22	169870197	<i>Coprinopsis cinerea</i> Okayama	MYRNFLYAA LSIVARSOLV GQTTTETHP MWQSCTAKG SCTCSDNKA CASNCAYDGA DYKGTGYITA SGNISLQKFI TKGSYSTHN SRITLMSDT AYQMFEPDGN KEFTFDVDS GLPGENGNAL YFVSMDDGG LKKYSGNKAG AKYTGKCD A QCPDLKPN SEGNVCKWP NDNANAGVDF GHSSCAEMD IWEANSLISTA VTPHACSTIE OTRODGDDG GRYSADRTAG VODPDGCDEN AYRMGEVKFY GKANTVDTISK KETVVTOTFG VMSLWDDHYS NMJWLDSTP TDKNPDTDLG SGRGSCDVKS GAPADESKS PDAVITYSNI KRGPLNSTDY
SEQ ID NO : 23	3913806	<i>Agaricus bisporus</i>	MLGKIAIASL SFLATAKGQO VGREVAENHP RLFWORCTR GGCOTVNSQO VVLDANWBWL HYTDGYTNCY TGNSPNSNSVC SDGTCQAEC ALEGANQQT YGTTSGNSL TMKFLITRSQG TNVGGRTVLM ENENRYQMN LLNKRAFTEDV DVSPVPGCGN GALYFVMDA DCGMSSOPNN RAGAKYGTGX CDSCOPRDIK FLDGVAMISVG WEPSEIDNSNA GRGYGICCA EMPIWEANM SNAVTPHCR T QNDGEGYRC EGRVTCNPYR EGICLDPDGC YNPFMNGKD FYGGKTDI NEFRNTVYOF ITHDNTDTGT LVDTRLYVO DGRVIANPPT NPGLMPAHD SITEBQCTDQ KNLFQDYSSR ARDQGLAHLNG RSLAKGHVLA LSIWNDHGAH MLWLDNSYPT DADPNKPGIA RGTCPTGTT PRETEQNHFD AQTFSNIKF GDIGSTFSGY
SEQ ID NO : 24	169611094	<i>Phaeosphaeria nodorum</i> SN15	MFPRSTLLAL SLTAVALGQO VGTNMAENHP SLMWORTSS GCONVNGKVUT LDANWRTHR INDFNTCYTG NEWDPSICPD GVTAAENCAL DGAYAGTYG VTISSGPAATL KFVIBESQKN IGSPLYNMAD DSNEYELPNLL NKEFFFDVY SKLFGCLNGA LYSEMAADG GMISSTNTAGA KYGTGKCDSQ CPDIKFDG EANSEGEGS PNDVAQGTGN FGACCGEMDI I WEANISSAY TPHCREPGL ORCEGENTCSV DGDKSYGPM TUDTNQPLTY VNGQITDNG DNGNLQELR YYVONGQVIO NSNNVNPQGD ID SENSISAEFC DQAKAEGFDE RSFDQRGGSU GMGSALORGW VLVLSTWDDH AVNMLWLDSD YPLDASPSP G1SRGTCSD SGKPDVEAN AGGTQVIVSN IKGDDINSTF NNINGGGNPN SPITTRPNSP AQTWMQGCGG QSWTGPFCQ SPSTCHIND FYSCCF
			MYRNIALASL SLRBAARAOQ AGTVTETHP SLSWKITCTGT GGTSCCTKAG KITLDANWRW THVTGTYTNIC YDGNSWNTIA CPDGATCIRN CAVGDADYSG TYGITISSNS LSISKEVTKGS NSANIGRTY LMESDTKYQ

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 25	3131	<i>Phanerochaete chrysosporium</i>	<p>ENLIGQEEFTF DVDTISKLPCG LNCALYFVEM AADGGIGKGN NKGAKYGTG YCDSQCPHDI KFINGKANVE      GWNPLADPN AGSGKIGACC PENDIWEANS ISTAYTIPHIC KGTFQLOCET DVSGGDGSN YFGQLCDRGCS      DFNSTRMGVK DFYCPGATLID TTKMVTWTO FLGSGSTLSE IKRFYVONGK VFKNSDSEALE GTGNNSITES      FCAAAQKTAFG DTNSFKTLGG LINENGASLAR GHVLYMSLWD DHAVNMLWLD STYPINSTKL GAQRTCAID      SGKPDVEKVN HPDDETVYFED IKGPIGSTF QOPS</p>
SEQ ID NO : 26	70991503	<i>Aspergillus fumigatus</i> Af293	<p>MVDIQIATEFL LLGIVGVAAQ ONGTYIPENH PLIATOSCTA SGGCTTSSSK TIVDANRWRMI HSTLGHITSCL      TANGDPTLC PDGIGTANCY ALIGVTSYST YGTTSSSAL RLOPQTGNTI GSROFLMADD THYTRFOLIN      QELADTDVDS KLPGLNGAL YFAMDAADGG KSKYKENRAG AKYGYGCGDS QCPRDVQFVIN GQANVQGNSNA      TSATGTTGSY GSCCTEILDW EAMSNAALT RCGSGNCTQ TGRCDAGNT FNSFRGLNTT      FLGAGMSVDT TKTETTYVQF ITSDNTSTGN LTEIREFVYQ NGNTYIPNSTV NYTGIAVNTS ITDPRECSQOK      KAFIEINTYFA QHGGLAQIUGQ ALRIGVWLAQF SISDDPAHNM LWLDSNEPPS ANPAPGVAF GRNCSTITSGNP      ADVGILNPSP YVSPFLNIEKG SIGTTERPA</p>
SEQ ID NO : 27	294196	<i>Phanerochaete chrysosporium</i>	<p>MHQALLESA LAYANAQV GTOQPETHPP LTWQKCTAAG SCQSQGSVY IDANWRWLHS TKDTTNCYTG      NTWNTSLCPD NESTRQNCL DGDYAQTG VTTSGSPLKL SFVTAANVG RLYTMODDET YOHNLNHAE      FTEDTDVSNL PCGNGALYF VAMDADGMS KYPNSKAGAK YGTGICDSCQ PRDLIKEING ANVEGWEPSS      SDKACVGHH GS CCPPEMDIW HPCDDAVSQT MSGDAGCQT YSERYAFTG DEDGCDNTPF      RMGNESFYGP GKIVDTKSFM TVYUQTATD GTDSGAVSEI KRLYVONGKV VSGNNTSTSDP      CTAQKAFGD EDIAKHGGL SNGKGALSEM VLIMSIWDDH HSNNMWLDST YPTDADPSKP GVARGTCBHG      AGDPNNEQS HDPSVTSN IKGPIGSTY EG</p>
SEQ ID NO : 28	18997123	<i>Thermoascus aurantiacus</i>	<p>MFRATLILAF TMAMVNGQO VGTNTAENHR LTLSOKCTS GCCSNLNTKI VLDANWRWLH STSGYTNCTY      GNQNTATLCP DGKTCANCA LDDADYCTY GITASGSSLK LQFUTGSNVG SRVYLMDTT HYQMFQLINQ      EPTFDVDMEN LPGCLNGALY LSZMDADGM AKYPTNKAGA KYETGKCDSQ CPDIKEING EANVEGHNT      SANATGNYG TCCTEMDIVE ANNDAAAYTP HFTCTNAQTR CSGSDCTRTD GLCDADGDFN NSFRGMDOFT      LGKGLTVDTS KPFIVTVQFI TNGTSGACTL TEIRRLYVON QKUIONSVK IPGDPVNSI TDNECSQOKT      AFQDNTYFAQ HGGKROVGEA LRTGMVLWS IHDYANML WLDSTXPTK DPSTPGVARG TCAITSGVPA      QIEAESPNAV VFNENIKFGD LNTTYTGTIV SSVSSSHSSS TISTSSSHSS STOPPTQPTCV TVPQWGCQG      IGYTSTTCIA SPYCHVLPN YYSQCY</p>
SEQ ID NO : 29	4204214	<i>Humicola grisea</i> var thermoidea	<p>MYORALLFSE FLAAARAHEA GTVTAENHPS LTWOCSSGG CCTTONGKVY IDANWRWHT TSGYTNCTY      NTWDSICPD DVTAQNCBL DGZDYSGTG VTTSGNARL NFVTOSSGN IGSRFLYLIQD DTYYQIKLL      GOEFEDDVY SNLFCGLNCA LYFVAMDADG NLISKYGNKA GAKYGTGCD SOCPRLDFPI NGOANVQGQ      PSANTDENAGV GNHGSSCAEM DPMVSSAET AVTPHPCDT QTWTMCGDDC GTTSPDRGCDP      NPYOCGNHSP YGPKIVDTS SKETVYTOFI TDGTGSGTL TEIRKFYVON GKVIPOSEST ISGVGTNSIT      TEYCTAQKAA FGDTGFFPHT GGJQKISQAL AQGMVLVMSL WDDHANMLW LDSTYPTDAD DTPTGVARGT      CPTTSGVPAD VEQONPNTSY TYSNIKYGP NSTFTAN</p>
SEQ ID NO : 30	34582632	<i>Trichoderma viride</i> (also known as <i>Hypocheaea rufa</i> )	<p>MOIKSYIQL AAALPLLSSV RAOAGAGITA ENHPRMTWK CSGPNCOTV QGEVVIDANW RWLHNNGNC      YEGNKWTSQC SSATDCAQRC ALIGANTQST YGASTSCEDSL TLKERVTHY GTNTGSRFVLY      TLMMNIEFAFD VDISKVECGT NSFALYFTAME EDDGMASYPS NRAGAKYGTG YCDAOCARDL KFIGKANIE      GWRPSTNDPN AGVOPMGACC AEIDVWSNA YAYAFTPHAC GSKNRYHICE TNNGGGTYSD DRFAGYCDAN      GCDVNPYRMG NKDGYGKGT VDTNTRKETVY SRFERNULSQ FEVODGRKIE VPPTPTWPSLP NSADITPELC      DAQFVFDDR NRFEETGGED ALNBALTIPM VLVMSIWDDH HSNNMWLDSS YPPEKAGLPG GDRGCPPTTS      GVPVAEQAQY PDAQVWSNI RPPIGSTVN V</p>
			<p>MYRKLAIVSA FLATAARAQSA CILQSETHPP LTWQKCSSGG TCTQQTGSVY IDANWRWTHA TNSSTNCTYGD      NTWSTSLCPD NETAKNCBL DGAYASTG VTTSGNSLTI GFTVOSAQRN VGFALYMAS DITYQEPTE      GNEFSEFDVY SQQLGCLNGA LYFVSMADG GIVSKYFTNTA GAKYGTGCD SQCPRLDKFPI NGOANVQGWE</p>

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 31	156712284	<i>Thermoascus aurantiacus</i>	PSSNNANTGI GHGSCCSM DIWEANSTSE ALTHPCTTV GOICEGDGC GGTYSDNRYG GTCDPDGODW DPYRILNTSF YGPSSFLD TIKLKLIVTQ FETSGAINYI VVQNEITFQQ PNAELGSTG NGNLDDYCTA EEAEGGSSP SDKGCLTOPK KATSGHMYLV MSLMDYYAN MLWLDSTYPT NEISSTPGAV RSCSCTSGV PAQVEQSPN AKVTSNPKF GPTGSTDPS GNPPGNAPP GTTSTRRPAT TTGSSPGTQ SHYGQCGGIG YSGPTVCASG TTCCQYLNPY SQCL
SEQ ID NO : 32	39977899	<i>Magnaporthe grisea</i> ( <i>oryzae</i> ) 70-15	MIRKTITLAA LYGIVRGQQA CSLTAETHPS LTWOKCSSGG SCITNAGSVT IDANWRWHT TSGYTNCTYGT NKWDSTICST NADGASKCCV DGANYQQTG ASTSGNALS QYTOSSGRN VGSRLYLIES ENRYQWENLL GNEFFEDVDA SKLGCLGLNGA VYFVSMADG GQSKYSGNKA GAKYGTGYCD SQCPDLKYI NGAANVBGWQ PSSGANSVG GMMSGCCAEM DIWEANSTI AYTPHP-CSNN AQHSCKGDDC GGYSSVYRA GDCDPDGDFP NSYRQGNRTF YGPSSNENVD SSKEKVTVTO PSSGGQLT TDGTGSGTL IKRFVYQNG VTFNQSOLIT GTVNNSVYTOQ YCDRQTAFFG DQNFENORG LRMGDAIAK GMVLYNSVMD DHHSQMLWLD STYPTTSTAP GAARGSCSTS SGKPSDVQSQ TPQATVYSN IKGPISGTF KSS
SEQ ID NO : 33	20986705	<i>Talaromyces emersonii</i>	MIRRALIIS SAILVKAQO AGTATAENHD PLTWOEQTAP QSCITQNGAV VLDANWRWHT DINGYTNCTY GTNTWDPYCP DDETAQNCQ AIDGADYETY GTVSSESSLK INFUTGSNVG SRATLLODS TYQIKULLNR EFSPEFDVDSN LPGCLGLNGALY FVZMDAGGV SKYPNNKAGA KYGTYGKCDSQ CPDLKEDGE EANVEGHQPS SNNANTGIGD HGSCCAEMDY WEANSTENAV TPHPCDTPGQ TMCSGDDCGG TYENDRYAGT CDPDGDFNP YRMGNTSPYQ PGKJLDTIKP FTWVTFQFLTD DCTDTGTLS E IKRFVTONSEN VIOPNNSDIS GTVNNSVYTOQ FCTAOKAOFG DTDFDSRQGG LANGAAMOO GMVLYNSLWD DYAAQNLWLD SDPTDAAPT TPGIARCTP TDSGPSDVS E SQSENSTVY SNKFGPINS TTTAS
SEQ ID NO : 34	22138843	<i>Aspergillus oryzae</i>	MHQRALLFA FWTAQVQQA GILTAETHPS LTWOKCAAGG TCTEOKGSVW LDWNWRWHT VDGSTNCTYGT NTWDTLCPD NESACSNCL DGADYEGTYG VTTSDGALTQ QFTVANANGS RLYMADDDE SYOTENULNN EFTFDYDASK LPGCLGLNGAVY FVSMADGGV AKYSTNKAGA KYGTYGKCDSQ CPDLKFING QVRKGWEPSD SDKNAGVGGH GSOPQMDLW EANISTATF PHPCDTTAQT MCEGDTGGT YSSERYAGTC DEDGCDENAY RMGNESPYGP SKLWDSSSPV TVTWTIPAD DTGSGALSE E IKRFVQGGV LANASNDV VTGNSITADP CTAOKAOFG DDTEAOHGGI QGQNLALSM VLTLSWDDH HSMMWLDSS YPEDADATAP GVARGTCEPH AGDPKEVQS QGSAITVYSN IKGPISGTF DAPA
SEQ ID NO : 35	55775695	<i>Penicillium chrysogenum</i>	MASTSFKIV KANLLAIAPL GAQAOQVGT STAEEHPSLT WOKTAGGSC TSQGKVVID SWRWRVHTG GYTNCYTGND WDRTLCPDYV TCNTNCAALDG ADYKGHTGYVT ASGSLLRLNF VTOASQRNKG SPYLMADDSS KYEMEQLLNQ EFTEDVDSN LPGLGALNY FVAMDEBGM ARYPTNKGAGA KYGTYGKCDQ CPDLKETING QANVNGWEPS SSDTNGGTGN YGSCCAEMDI WEANSTTAF THPHCDDPAQ TRCTGDSGG TYSSDRYGT CDPDGDFNP YRMQNOSEFG PSLKIVDTEP FTWVTFQFLTD DGSTTGTLSE IKRFVYQNGR VIPOSVSTIS AVTGENSITDS ECSSQKTAFK DUVFAKHGG MAGMGAGLAE GMVLYNSLWD DHAANMNLID STYPTSASST TPQARGSCD ISSBPSDVE ANHSNAVVY SNIKVQPLGS TFGSTDGSGG TTITKVTITV AVKTTTITGP STTGAHYAQ CGGQWNTGPT TCSPYPTCQD QDYYSQCL
SEQ ID NO : 36	171676762	<i>Podospora anserina</i>	MVSAKFAALA ALVVASAQQ VCSLTPESHP PLTWQRCASG GSCTNVAQSV TLDSNWRWHT TLOGSTNCTYS GNEDWTSICST TGTCIAQNC VIEAAYAATY GTTSGRQLN LKFTVTEGKYS TNVGSRTLM ENATKYOEN LLGNEEFEDV DVSNIGGCLN GALFYVSMDL DGGLAYSGN KAGAKYGTG CD:QCPRDIK FINGEANIEG

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 37	146350520	<i>Pleurotus</i> sp Florida	<pre> WNPSTNDVNA GAGRYGTCCS EMDIWEANNN ATAYTPHSCT LLDOSRCGEV SCGGYTSSDR YGGVCDPDGC DFTNSTRMGNK EFTYQGVTEF TIKRMTVTTQ FLRNAGELG EKIRYVGEV VVIFPNQNIS IT PFPVNQNIS IT QDWCMQAOKIA FGDDDDNTAK GGIRGMGIAL DRPMVLMISI WNDHAHMLW LDSTYPVDAA GRPGAEKGAC PTTSGVPEV EAEPNNSVA FSNIKFGPIG STENSSSTNP NPISSTATT PTSTRVSSTS TAQQTTSAP GGTVRWGQC GGQGTYGPTQ CYZPYTCVVS NQWYSQCL </pre>
SEQ ID NO : 38	37732123	<i>Gibberella zeae</i>	<pre> MPEYLALVSP SPLSVLVLAQO VEGUTAETHP QLTVOQCTRQ GSCTTQORSV VILDGNMRWLH STSGSNINCYT GNTWDTSCLCP DAAFSRNRNA LQADSYSTY GITSGRLAATL LKFVTHGPYS TNIGSRVLL ADDSHYOMEN LNKNKFTEDV DVSOULPCGN GALYFSONDA DGCTGFEPNN KAGAKYGTGY CDSCQCPHDITK FTNGEANVOG WQPSENDNSNA GKGQYGSVCA DCGGSCCA ASAYTPHCTP VTTPLRCQGN GVCVDKDGDFP NSFRGDKNF LGPKTNTNNT SKTYYVTOFL TSDNTTSGTL SEIRLYVQN GRYVLSKVN IPGMASLUDS ITESFCSTQK TVFBDTNSPA SKGLRANGN AFDKGAVLVL SIWSDHEAKM KSASAPGVAR GTCAATSGEP KDVEQSOSPA QVTFNSNIKYG DIGSTYEN </pre>
SEQ ID NO : 39	156055188	<i>Sclerotinia sclerotiorum</i> 1980	<pre> MYRAJATASA LIAAVRAQV CSTLQEKSPLS LNWSKCCTSSG CSNYKGSVTI DANWRWTHQV SGSTNCYTGN KWDITSEVTSG KVCEERCCU GAYAYSTGI TSQDQSLIS FVTRCPYSTN IGSTYLMED ENTYQMFQLL GNEFFEDVDV SNIGCGLNGA LYFVSMADG GKAKYFGNKA GAKYGTGYCD AQCPRDVKFI NGQANSQDGWQ PSDSDVNGGI GNLOLTCPPM DIWEANSLIST AYTYPHPCTKL TQHSGCTGSC GGTYSNDEYQ GTCDADGDFP NSYRGKNTF YGPSSGENVD ITKKVTVTYO PHKGENGRLS EITRILYVONG KVIANSESKI AGVPGENSLTA DFCTIQKVF NDPIPDFTRKG AWGMSDALLE AFMVLVNMWL HDHHENNMWL DSTYPTDSTK LGSQRGSCST SSGVADLEK NVPNISKVAFS NIKFGPIGST YKSDGTTPTN PTNFSEPSNT ANPNPGTVDO WGQCGGSSNYS GPTACKSGFT CKKINDFTSQ CQ </pre>
SEQ ID NO : 40	453224	<i>Phanerochaete chrysosporium</i>	<pre> MYSAAVLFATE SFLLGAGACQ VGTLKTESHP PLTIQKAAG GCTCDBADSV VLDANWRWLH STSGSTNCYT GNTWDTSCLCP DAATCTANCA EDGADYEGTY GITSSGDSLK LSIVTGSNVG SRTYLMDFSET TYKEFALLN EFTFTYDVSK LPCCINGALY FVNDMADGQM SKYPTKNGA KYGTYKCDAO CPQDMKFKSG DENPYRNGNT SNSANSGTGN TGSCSCEFDV TQHSGCTGSC NTGUCDADCGC SDISVGNSN ITDDFCPAOK TFYGGKTID TTKEFSTVTOU FTTDDGTSQ TLTEIRFVV QDDVYEEPN SDISVGNSN ITDDFCPAOK TAGFTDFTFS QKGGMAMCK KJZDGMVILV SIMDDYNNNM WLSDLSDYPTI KDASTPGYSR GSCATTSGP ATVEASGSA YTTESSIKPG PIGSTFKA PA DSSSPYVASS SPAAYAVVS TSSAOAVSH PAVSSSQAV STPEAVSSAP EVPSSSAQ SYVPTSTPKP CSKVQSSTL ATSVAAPATI ATSVAATTS AASSSGSYPL YGNCTCGKTC SECNCVQNP WTQSQCVSS </pre>
SEQ ID NO : 41	50402144	<i>Trichoderma reesei</i>	<pre> MFRAAALLFA TCLAIVSGQ AGNTAAENHP QLOSOOCCTP GCKPLSTKV VLDONWRWTH STSGYTCYT GNEMWDTLCP DGKTCANCA LDADSYSTY GITSTGALT LKFVTHGPYS VLDONWRWTH STSGYTCYT EFTFTYDVSK LPCCINGALY LSAMDADGM SKYPTKNGA KYGTYKCDAO CPKDIKFING EANVGNTEF GSNTGTSYV TCCEBMDTWE ANNDAAHTP HCTTGTQTR CSGDDAART GLCDGDCDF NSFRGDKTF LGKMTVDTK KPFTVUTOPU TNIDNTSTCTL SEIRRATYNG GKVIOSVAN IPGYDPVNSI TDNFCAQKT AGDINWFAQ KGKGKQMGEA LGGMVNLALS IWDHHANMLW LDSDYPTDK DPSSAEGVARG TCAITSEVP DVEQSYPNSO VVFNENIKGED IGSTFSGTSS PNPPGGSSTTS SVTTSPTPP PTGTTVPOKG QCGGIGTSGS TTCAСПТЧ VLNYYSCQY </pre>
SEQ ID NO : 42			<pre> MYRKAVISA FLATPARAQSA CTIQUSETHPP LIWQKCSSGG TCTOOGTSVW IDANWRWTHA TNISSTNCYDG NTWSSTLCPD NETAKNCCU DGAAYA-STYQ VTTSGENSLSI GEVTOSAQKN VGRARLYMAS DTYYQEFLL GNEFFEDVDV SQQEGLNGA LYFVSMADG GVSKYPTINTA GAKYGTGYCD SQCPRDVKFI NGQANVNGWE PSSTNANTGI GGHSGCCSEM DIWEANSTSE ALTPHPCTVY GOEICEGDGC GGTYSDRHYG GMCDPDGCDW NPYRUGNTSF YGPSSSFTELD TPKLTVVTTQ FETSGALINRY YVQNGEVTFQO PNABLGTSVG NEISSTPGAV RGSCTSSSGV EEAEFGGSSP SDKGGLTQFK KATISGGMVLV MSLMDYYAN MLMLDSTYPT NEWSTPGAV TCAITSEVP PAQVSQSPN AKVTFNSNPK GPGSTNPS GENPPGNGR TTTRPATT TGSSPPTPP PTGTTVPOKG QCGGIGTSGS SGPTVCASGT TCQVLNPPYS QCL </pre>

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 42	115397177	<i>Aspergillus terreus</i> NIH624	MPSDTIYKK LLLIASLFLA SQOQVQGTSK AIVHPSLTWQ TCTSGSSCIT VNGKVUVDAN WFWHVHNDGY NNCYGTNTWD TFLCPDDETC ASNCALEAD YSGTYGVTT SNSLRLNFTV QASOKNTGSR LYLMEDISTY KMFKLINQEF TFDYDVSNLP CGINGAVIFV SMADGSMAK YPAKAGAKY GITGCDSCQP RULKFINGMA NVEGEPSEN DANGTGNGH SCGAEMDWE ANSISIATYP HPCTDPGOM CTCDSGGTY SSDRYGETCD PDGCENSYR QGNKTFYGFQ MTDQFLTSD TASFGLTSD NSNMWLDN YPTDKPSTP ENSESTWSVG SGNSTTAYC NAQFTLFTD DUTVKHGMG GNGAAALAEGL VVVLSLWDDH NSNMWLDN YPTDKPSTP GVARSQDIS SGDPRVEAN DANAYVYNSN IKGVTGSTF SGSGGSSS STIAKSITTT TSATKITT QKONDYYSQCL L
SEQ ID NO : 43	154312003	<i>Botryotinia fuckeliana</i> B05-10	MISKVLAFTS LLAAARAQQA GTLTTETHP LSVSQTASG CTTSQAQIVV DANWRHLST TGSTNCYTGN TWDKTLCPDG ATCAANCALD GAIYSGYGI TTSQNSIKLN FVTQKANTNV GSRYLMLAG STTQYQMLKL INQETTEDV VSNUPLCGUNG ALYFAAMAD GCLSREPTNK AGAKYGTGYC DAQCPDQIKF INGVANSYGW TPSSDVNAG AGQTGSCCE MDIWEANKIS AAYTPHICPNN DTQTRCTGQD CGIGARYSSL CDAGDCDENS YRGNTSFYQ AGHTVNTNPKY FTWVTOFTIN DCTASATGLKE IRRETVYQNG VIPSNTSIA GVPGNSTIDS FCAAQATARG DTNEBAPATCGG LATMISKALAK GNVLYNSIW DHTANMILWD APTPATKSPS AFGVTRGSCS ATSGMPDV E ANSGSSSTVF SNIKWGPINS TTGSGAAPS VPGTTVSSA PASTATSGAG GYAKYAQCGG SGYSGATACY SGSTPCVALNP YYSQCC
SEQ ID NO : 44	49333365	<i>Volvariella volvacea</i>	MFPAAFLAFP SLRAFYVGQ VGHQLAETHP RLTLWQCTRNS GGCQOTOSNGA IVILDANWWRV HNVGGYTNCY TGNTNTNSLC PGATCACNC ALDGANTYST YGTTTSGNAL TLKFTVQSEQ KNIGSRVYL ESDTKYQLEN PLNOQETEDV DVSOPLPGCN GAVYSGACCN DCGMSKFPNN AGAKYGTGYC CDSCGSPRDIK FINGEANEVQG WOPSNNDTNA CTGMYGACCN EMVWEANSI STAYTHPACT QOGLVRCSTG ACGGSNRYC STCDPDGCD NSFRGDKSPF YGPHTVNTQ QKETVTVTOFL TNINSSGTL REIRLYVQN GRTVQNSKVN TPGMPSTMDV VTTERCNAQK TAFNDTEFSQ QKCGMAMMSE ALRRGMVLV LWDSDHAAMM LWLDSNYPTD RPASQPGYAR GTCPFSSGKP SDVENSTANS OTYTSN1KFG DIGSTYSA
SEQ ID NO : 45	729650	<i>Penicillium janthinellum</i>	MKGSI SYQIY KGALLSALL NSYQAOQVT LTAEHTPALT WSKCTAGKCS QVGSVVDA NWPXVHSTSG STNCYGTNTW DATLCPDDT C3ANCAVDA RQHRLRVTS GNSLRINFTV TASOKNTGSR LYLLENDTY QKENLILQEF TFDYDVSNLP CGINGALYFV DMDADGMAK YPTNKAGAKY GITGCDSCQP RULKFINGOA NDGWPSPSKN DUNSGTGNHG SCGAEMDWE ANSISIATYP HPCTDPGOM CTCDSGGTY STDYGGTCD PDGCENSYR MGTVNFYGFQ ETIDTKPFT VTOFLTNDG TSTGLTSEIK RFYQGGERVY CNPOSTIVGV SGNSITDWSW NAQKSAGDFT NEPSKHGMA GNGAGLADM VVVLSLWDDH ASDMILWLDST YPTNATSTP GAKRGTCDIS RRPHTVESTY PHVYVTSNI KTGPNLSTFT GTTSSSSSTT TTTSKSTSTS SSSKTTTWT TTTTSGSSG TGADWAQCG GROWTGPTC VSPYPTKQN DWYSCQL
SEQ ID NO : 46	146424871	<i>Pleurotus</i> sp Florida	MFRTAALTAF TLAVVLGQ VGTLTAENHP ALSIQCTAS GCTTQKSVV LDNMRWTHS LPVHTNCYTG NAWDSLCPD PTCAATCAI DGADYSGTG ITTSQNLTL RFVTINGPSK NIGSRVYL DADHYKMFDL KNQEFTFDV MSGPFCGUNG ALFSEMPAD GGPAHDFKW INGEANLILW DAQCPHDIKW YMEFLIANE SASATDANAG NGRGACCAE MDIWEANSEA TAYTPHVCRD EGLYRCSGTE CGDGDNRVGG VCDKGCDFN SYRMDRNFL GRGETIDTJK KI TVTVOPIIT DNTTSSGNLV EIRRVYQDG VTVQNSFSTF PSLSQNSIS DDFCAQKTL EGDNQYXNTH GGEKMKDAM ANGMVLMISL WSDHAHAMLN LDSYPLDKS PSEPGVVERGA CATTGDPDD VVAHPMASV TESNIKTP1 GSTYGGSTPP VSSGNTSAPP VSTTSSGPT TPTGPTGTPVW KNGQGCGNGY SGPTCVAGS TCYTSNWDY QCL
SEQ ID NO : 47	67538012	<i>Aspergillus nidulans</i> FGSC A4	MYQRALLEA LLSTVSRAQQA GTAQEEVHPS LTWORCEASG SCTEVAGSVV LDSMRWTHS VDGTYNCYTG NEWDTLCPD NESAQNCAV DGADEAYTF ITSGNSDSTL KFVTSNSVGS RVFLMEDDET YMEFLIANE FTEDDYDVSNL PCGNGALYF TSNDADGSL KYEGNTAGAK YGTGCDSCQ PRD1KFINGL ENVEGMPEPSD SDANAVGGM GTCPEMDIW EANSISIATYP MCEGDSGGT YSDDRYGGTC DEDGCDENSY RMGNTSFYGP GAIIDTSSKF TVVQFIADG GSLSEIKRFY VQNGEVIPNS ESNISGVENI SITSEFCCTAQ

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 48	62006162	<i>Fusarium poae</i>	KTAFCEDEIF AQQGGLSANG DASAMYILIL SWDDHHSSM MWLSSYPTD ADPSQPVYAR GTCEQGAGDP DVVESEHADA SVTFNSNIRFG PIGSTP
SEQ ID NO : 49	146424873	<i>Pleurotus</i> sp <i>Florida</i>	MYRAATASA LIAAVRAAQV CSLTTETKPA LTWSKCCTSSG CSVNGSVTI DAWRWTQW SCSTNCHTGN KWDITSYCTSG KVCEBKCCVD GADYASTYGI TSGNGNLISL FVTGSGYTN IGBTYLYMED ENTYOMFOLL GNEFTFDVDV SNICGGNLGA LIVCFSMDAIG ACPRDVKFI NGQANSIDWE PSKSPTYNGGI GUICGCCPRM DIWEANSTST AYTPHPTCKL TQAHATGDSC GGTYSNDEYG CICDADGDF NAYROENKTF YGPSSGFNDV TTAKVTVUTQ RHGSNGNRLS ETIUYVQNG KVIANSESKI AGNPGSLLTS DECTQKVKV GDIDIFDAKKG ANVGMDSALE APMVLYNSWL HDHHENNMLM DSTYPTDSTA LGSORGCST SSGVADLEK NVPNISKVARS NIKGP1GST YNKEGTQQPQ TPNTNPNPNTN PTNGPTGTVQW GCGGGTNYSG PTACRSPFTC KKINDFYSQ C
SEQ ID NO : 50	295937	<i>Trichoderma viride</i>	MERTIALTAF TLAAVILGQO VENLAARENHP ALSIQOCTAS GCTTQOKSTV LDNNWRWHTS TAGATNICYTG NAWDSSLCPN PTTGNCNCAI DGDYSGTYG ITTSGNLSL RGVYKNGVW DADHYKLFNL KNOEFEDVDP MSGHPGCHNG ITLYSENAAD GGKAAGHTGNV AFKVYGTGYC DAQCPDHDKW INGEANLILDW SGSATDPNAG NGRGACCNE MDIWEANSEA TAYPHVCRD EGLYTRCSGTE CGDDDNRYTG YCDKDGCDFN SYRMGDKNFL GRKTIDTTFK KITVVTQFIT DDNTPIGNYL FIRRYVYQDG VTONSFSTP PSLSQYNISI DDFCAQKTL FGDIQYNTN GGPKMGBDSL ANGMVLMISL WSDRAHAMLN LDSDYPLDKS PSEPGVSRGA CATTIGDPPD VVANHQNASY TESNMKYGPI GSTYGGSTPP VTSNTSSGPT VTSTTSSGPT TPTGPTGTVP KWGQCGGIGY SGPTSCVAGS TCYTSNEWY QCL
SEQ ID NO : 51	6179889 #	<i>Alternaria alternata</i>	MYOKLALISA FLATARAQSA CTLOAETHPP LTWQKCSSGG TCTQGTGSVY TDNSTMCYDG NTWSTLICPD NETAKNCIL DZQAYASTYV VTTGSAISL VTCYTSQAVN VCGYLINMS DTYQEFIL GNEFRFDVDP SGLCPGHLA LYTFSMDAIG GTVKYPINTA GAKYGTGYC SQCPDLIFKI NGQANVYGWIE PSSNANTGJ GGHGSCCSEM DIWEANSTSE ALTPHPCFTV GOEICEGDS C GTTYSEDRYG GTCDPDGCDW NPYRUGNTSF YGPSSFTLUD TTKLITVUTQ PFTSGAINRY YVQNCVTFQO PNABLGDSG NSLDDDCFAA EEAEGGSSFE SDKGSLTQK VDQKMTYAN MSLWDDYYAN MLWLDSTPYV DETSPTGAV RGSSSTSSECV PAOLSNNSPN AKVYNSNPK GPITGSTGNS GPITGSTGNS GTTGPRTAPS TGSSGPPTQI HYQCGGIGY IGPTCAGSGS TCQVLNPYI QCL
SEQ ID NO : 52	119483864	<i>Neosartorya fischeri</i>	MTWQSCTAGK SCINNKNGKV IDEANWLRHK KEGYDNCYTG NEWDATACPD NKACAANCVA DGADYSGTYG ITAGENSLSKL KFTIKGYSST NIGCSRTRYLWK DDTTYMFKF TGNOEFTFDV DVSNLPGCFN GALYFVMDA DGGLKKYSTN KAGKKGKTY CDZQCPDLK FINGEGRVEG WKPSNDANA GVGHGHSCCA EMDIWEANSV STAVPHSCS TIEOSRCDGD GCGTYSADR YAGVCPDGC DFNSTRMGVK DFYGKGVIVD TSKKFTVYTO FIGGDAMEI KRFTVQNGKT TQGASAVPQI DQQAQAVFQD CQKAVAFGD VTFKKGMM ANMAKALANG MVLVNLMDD HYSNMLWDS TYETDKNPDFT DLGTGERCECE TSSTGPADVE SOHADATVYV SNIKFGPLNS TFG
SEQ ID NO : 53	85083281	<i>Neurospora crassa</i> OR74A	MASAISFOVY RSALILSAFL PSITOAOQIG TYTTEPHSM TWETCTSGGS CATNGSYVM DANWRWYHOV NTHYOMFKLI NOETTFDVFV SNIQGUNGA LYFVIMDEDG GVSKYPNKVA GAQYGVGFCD SQCPRDUKFI QGQANVEGWT PSSNENTGFL GHYNGSCAEL DIWEANSISQ ALTPHPCDTA INTMCTGDAC GETYSSIDEYA GTCDPDGCDP NPYRNGNTF YGKGTIDTN SPFTVYTOFI TDDGMDTGTL SEFRRYYON GVTYAAQPSD ISGITGNAIN ADYCAENTV FDDGTFARKH GFPSANSEAM STGMVLVMSL LDSTYPTMAS SSTPGEAVRS CSTSUGPAT ISESEPSDVY TYSNIVKGP1 GSTRSGSGS GSSESGSSGS ASTSTSTK TAATSTSTAV AQHQSQCGQ DWVGTPTCVS PYTCQVONAY YSQL
SEQ ID NO : 54			MKAYFEYLVA ALPIGLLATA QVGKQFETT HPLSMRKCT GRANCNTVA EVYIDSNPRW LHDSGKNRYC DGNKNTSAC SATVCAASKQ LIQCANGTYT GAITSSEDALT LKEYTKHEYG TNIGSRFYM NGASKYQMF LMNNBFAFDV DLSTVEGEN AALYFVAMEE DGGMASYSN KAGAKYGTGY CDAACARDIK FVGGKANIEG WTPSTMNDANA GVGGYGGCCA ELLWESNAH SFAFTPHACK TINKHVVCERD NGGTYSEDVR FAGLCDANGC

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 54	3913803	Cryphonectria parasitica	DYNPYRMGNT DFYCGKTKVD TSKKFTVYSR FEENKLTOFF VONGKIEIP GPKMDGIPSD NANITPEFCG AQFQAFGDRD RFQVGGFAQ INSGRSLPWF LMWSIMDDH ANMILWDSVI PPEREGQQGA ARGDCPQSG VPAEVESQYA NSKVYVSINR FGVGSTNV
SEQ ID NO : 55	60729633	<i>Corticium rollei</i>	MFSKPALTGS LLAGAVNAQG VETQOETEHP QMTWOSCTSP SCSTNQEV VIDSNNWRVH DKDGYYNCYT GNTWMTTLLCP DDKTCAAANCY LDGADYSSRT GLTSGNALS LQFTQSSGK NIGSRTLYME SSKYHFLDL IENEFAPFDV D LSKIPCGHNG ALFVFTYMDA GCMAKYSTNT AGAEFGTGYC DSQCPDRIKE INQGUNVEGW TPSTNDANAG VGGIUSCCSB MDWEANSMID MAYTPHACET AAQHSCNADE CGGYISSSSY AGDCDPPGCD WNPPFMNGKD FYGGGDTVDT SQKFTVYTF HGSGSSETEI SOYTIOGGK IQQPNTSTPPT LGYNTSTTD FCKAQKEFEFN DTDPFSEKGG LAGMAGNAID GMVLYMSLWD DHYANILWLD STYVVDADAS SPGKQRGTC TTSGYPADVE SSDASATIVY SNIKFGPTIGA TY
SEQ ID NO : 56	39971383	<i>Magnaporthe grisea</i> 70-15	MFPAAHLLSP TLIAAVASQGQ IGNNTAAEHP SILTVSQCTTS GGCSTSTQSI VLDANWRMLH STSGYTYCCT LNQWNSDLCP DSDPDCATGCA LDGASYSDT GISTDGNAT LNFTYTGQST INGSRVPLIS DTHYQTFSL LNKEFSDV D ASNLGGCING AVTFYQVNDAD GGLSKYFSSNK AGAQYGTGYC DSQCPDRIKE INGEANULLDW NATSANSGTG SYGSGSCPEND IWFANKTAAGA YTHPHPCSVSG QTRCYGTSGC AGSRYDYG C DKDGCDENSW RMGNETFLGP GMTIDTNKPF TIWTQFTIDD NTANGTLESEI RRLYVQGGTY IONSVANOPN IPKVNSITDS FCTAQTTERF DODFTGTTG LSMGKAMSD MLVLMSTWDD YDAEMILWDS NYPSGSAST PGISRGPSA TSGLIAVTEES QQASASVTS NIKWGDIGST YSGSGSSGSS SSSSSAAAATATSSAAAAT GSPVPAYGQC GGQSYTGSTT CASPVYCKVS NAYYSQCLPA
SEQ ID NO : 57	39973029	<i>Magnaporthe grisea</i> 70-15	MKRALCASL LLAABAQW GINPEPEVHPK MTKWKGSSG SCSTNQEVV IGGYENCYSG NKFWSYCSCN ADGCTKACM GAKYQETGKU STSGDALTIK FQOQNISVGGN UGSPRMWLNH ANKYQMFTLK NNEFADFVDL SSVEGMNSA LYTFPMKEDG GMSTEPRNKA GAKYGTGYC ACCARDLKF VGGKANTEGW PSSTSSASSGI GAQGCCACAE DIVESMRNA AFTPHPCENN EYHVCTEPNC GGTYADDRYQ GGCDAANGDY NPYRIGCNPDF YGPERTIDIN RKTVIERSEI NMRNYOILMO DGVAIRIHPG KFDGLEGETG ELNEQFCCTDQ FTVFDERNF NEVGCGWSKLN AAEYIPIWLV NMISWSDHFAN MLWLDSTYP EKACQGPSAR GPCPADGSDP NGVYVQYNA KVIVSNVRQG PIQSTQYD
SEQ ID NO : 58	1170141	<i>Fusarium oxysporum</i>	MOLTAGVFL GALMGAAAQ OVTQOTAENH PKMTWKCTG KASCTVNGE VVIDANWRWL HDASSKNCYD GNRWTDSCTR ASDAAKCSL TSQSGDALLK ASTSGDALLK KFVTRHDGT NIGSRFLMN GASKYQMSL LGNEFADFVD LSTTECGLNS ALYFVMAED GCMKSYSSNK AGAKYGTGYC DACCARDLKF VGGKANTEGW KPSSDNDANAG VGPFGACCE IDWESNAHA PAFTPHPCD MKYHVQCDN CGGYSDDEF AGKCDANGCD INPVRLGNTD FYGGKTKVD SKRFTVYTF ERDALTOFFY QNNKREIDMPS PALEGFLPTG ALTAEYCTNIV FNVFGDRNRF DEVGGWSQLO QALVLPWLV MSIWDDHYN MLWLDVYPP DKEGSPGAR GDCPQDGSVP SEVEEQIPGA TVVWNSNIREG PIQSTQYD
SEQ ID NO : 59	121710012	<i>Aspergillus clavatus</i> NRRL 1	MYRITASATA LIAARAQO CSLNNTETKPA LTWSKCTSSG CSYDKGSVVI DANWRTHOT SGSTMNCYTG GNRWTDSCTR ASDAAKCSL TSQSGDALLK ASTSGDALLK KFVTRHDGT NIGSRFLMN GASKYQMSL GNEFADFVDV SGICGGLNSA PHVFSMDEDG GAKYGTGYC AOCPRDVIFP IGVANSTGK PSDSDFVNVAGV GNLTGTCPEM DIWEANSLIST AFTPHPC-TKL TQHSCCTGDSC GGYSSDRYQ GTCDAADGCD NAVYQGNKTF YGPOSENNUID TPKMTVYTO FHKGSNGRLLS EITRLUYQONG KVIANSEEKI ACNPGSSEUTS DFCSTQKSVF GDIDDFSKGG GNMGMSDALS AMVLVMSLW HDHASNMLW DSVTPDSTSX VGSQRGSCAT TSGKTSUDLER DVPNISKVSPS NIKKEFPIGST YKSDGUTTPNP PASSSTTGSS TPTNPPGASV DWGQCGSN YSGPTCKSP FTCKKINDFY SQCG
SEQ ID NO : 60			MYQRALLFSA LATVASSAQV GTCRKAEVHPA LTWQKCTAAG SCTDQKGSVW IDANWRWLHS TBTDTNCYTG NEMNAELCPD NEAGAKNCL DGADYSCTYQ VTAGDSSLKL NEWTSFANVGS RLVLIMDEDT YQMFNLNN PTFDVDSNL PCGNGALPF VSNDADEGSL KYPGNZAGAK YGTGYCDQSC PRDJKFNGEE ANVEGWPSD DNKNAIAGVGGY GSCCPMEMDIW EANSISTAYT PHPDGMEQT RCDGNDGGT YSSTRYAATC DPDGCDFTNSP

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 60	17902580	<i>Penicillium</i> <i>funiculosum</i>	RMGNBSFYGP GGIVDTKSPV TYWTOFVTAG GDSGALKEI RRVYVQGGKV IGSASNVAG VEGDSITSDP CTAQKAFGD EDIFSRKGGL EDKSGLNLKNN ALIVS1WDDH ASMMWLDST YPVADADATP GVARGTCBHG LGDPETVESQ HDPSAVTFRN IKGPIGSTY KSV
SEQ ID NO : 61	13462226	<i>Humicola grisea</i> var <i>thermoidea</i>	MSALISFMNY KSALILGSLL ATAGAOOIGT YIAETHPSLS WSTCKSGSC TTNSGATLTD ANWRWVHGVN TSTNCYTGTN WNTAICDIDTA CGNGCALDVG ADYSGTVGTT MNGSLRNLF VIGSNVGSFT LMADNTHYQ JEDLJNQEFT FTVVSNLPC GLNGALJEVFT MDADGVSKY PNNAGAQYV VGYCDSSOPR DIKFIAQGN VEGWNPSTNN SNTGIGNHOS CCABLDWEA NSISEALTHP PCDTGTLVC TADDCGGTYS SNRYAGTCDP DGCDPNPYRL GIVTFYGSCK TWTTKPTV VTOFYVDDGT SSGSUSEIRR YYTQNGVYVIP QPSSKLGIS GNVINSDRCA AELSAFGETA SFTNHGGIURN MGSALEAGMV LVMSLMDYS VNMWLWDSTY PANETGTPGA ARGSCPTTSG NPKVVEQSOG SSTVYFESDIK VGPENSTESG GTSTGGSTT TAAGTTSKKA STTSTSSTT GTGVLAHMWQ CGGGWGTGT TCAGTCTV VNPYVSCCL
SEQ ID NO : 62	156712282	<i>Chaetomium</i> <i>thermophilum</i>	MRTAKFATLA ALVSSAAAO ACSSLTERHP SLSWNKCTAG QQCQTVQASI TLDSSNWRTH QVSGSTNCYT GNKNTSIC TDAKSCAQNC VDGAQADSYSTY GITTNGDSLS LKFVTKHQHG TNVGSRTYLM DGEDKYQTEF LLGNEFTEDV DVNTIGCEN GALIYFSNDA DCGMSKXSN KAGAKYGTGY CDACCPRD1K FINGEANIEG WTGSTNDPNA GAGGYGTCCS ENDIWEANNN ATAAFPHPCT IIGOSRCEDG SCGGTSYNER YAGVCDPDGC DFNSTROQNK TFYKQGMITD TTJKITVTO FLKDANGDLG EIKTRVYQDG KLTNPSESTI PGVEGNITQ DWCDRQKVAF GDIDENRKG GMKOMGHALA GPMLVLYMSW DDHASNMWL DSTFPVDAAG KPGAEARGACP TTSGPAAVE AEANNSNTVF SNIRGPQIGS TVAGLPAGIN GNNGNGNPPP PTITTSSAPA TTITASAGPK AGRMQOCGGI GFTOPTCCE PYTCTKLNDW YSQL
SEQ ID NO : 63	169768818	<i>Aspergillus oryzae</i> RIB40	MYKVKPANLA ALVZGASAQQ ACSLTAAENHD SLTWRKTSG CSGCTVNGAV TIDANWRTH TVSGSTNCYT GNQNTSCLT DGRSCAQTC CVDGADSYSTY GITTNGDSLN LKFVTKHQHG TNVGSRTYLM ENDTKYONFE LLGNEFTEDV DVNLNGCEN GALIYFSNDA DCGMSKXSN KAGAKYGTGY CDACCPRD1K FINGEANIEG WTPSTNDANA GFGYGYGSCS ENDIWEANNN ATAAFPHPCT TCGTYESSDR YAGVCDPDGC DFNARQDKX SNTIDENRG GMQAMSKALA GPMLVLYMSW DDHTANMWL DSTPIDIQG ARGAERGACP EYCDOKVAF SNTIDENRG GMQAMSKALA GPMLVLYMSW DDHTANMWL DSTPIDIQG ARGAERGACP TTSGPAAE AQVENSNTVF SNIRGPQIGS TVPGLDGSNP GNPTTIVVPP ASTTSRPTS STSSPVSTPT GQPGCCTOK WGCGGGIGT GCTCNVAGTT CTQNPWYSQ CL
SEQ ID NO : 64	46241270	<i>Gibberella pulicaris</i>	MASLISLISKIC RNALILSSVL STAAGOOGT YOTETHPSMT WOTCNGGSC STNGSVYLD ANWRWVHOTG SSSNCYTGNK WDTSYCSSTD AZQKCALDG ADYNTYGTIT TSCTEVRLNF VTNNSNGKVN GSRVYMMAD THYEYTKLLN QEFFDVDS KLFGLNLGAL YFVMDADGG VSKYPNKAG AKYTGTYCDS OCPRDJKFIO GQANTBGWVS STIMANTGTG NHGCCAELD IWEANS1ST LTPHPCDPT NLCTGTGAG GTYSSDRYSG TCDPQCDEN PYRGNTTYP GPKTNTDNK PIUTVTOFIT DGDGSSGTLT EIKRFYVODG VTYPOPAEADV SGLSGATINS EYCTAENTLF EGGSFAKH GLAGMGBAMS TMWLVMSLW DDYANMMLW DSNYPNTNEST SKPGTARGTC STSSGVPSEV EASNPNSAYA YSNIKVGPIG STPKS
SEQ ID NO : 65	49333363	<i>Volvariella volvacea</i>	MYRATAATASA LIAAVRAQGY CSLTPETKPA LWSWKCTSSG CSYQGSVTI DANWRWTHOL SGSTNCYTGN KWDTSCTSG KVCEBKCCUD GAEYASTGYI TSQNGQJLSL FVTKGAYGTN IGSPTYLMED ENTYQMQLL GNEFFDVDS SNTGCGLNGA LYEVSMADG GKAQYGTGKNA GAKYGTGKNA AOPRDUVFI NGOANSQGWQ PSKSDVNAGI GNMGTCCPEN DIWEANS1ST AYTPHBCTKL TQHSCGTGSC GGTYSNDEYG GTCDADGCDP NAVRGNKTFR YGPOSGRENVD TTJKVUTVTO FHKGSNGRLS ELTRLYQONG KVIANSEKU AGVPGSSLT EFCITQKVKF GDTIDFAKKG AWGMSDASE AMVILVMSLW HDHHSNMMLW DSTYPTDSTK LGAQORGCSST SSGVPADELEX NVPNSKVAFS NTKEFGPIGST YKEGVPEPTN PTPNPTNPTN TNPSTVQDMA QCGGTNTYSGP TACKSPFTCK KINDFYSQCL
			MFPKSSLLVL SELTAYAQO VGTQTAEVHP SLNWARCTSS GCTNVAQSVT LDANWRWLIH TSGYTNCTYQ NSWNTLCPD GATCAQNCAL DGANYQSTCG ITTSGENALTI KFVTOQEQRN IGSRVLYMAS ESRYEMFGLL

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 66	46395332	<i>Trpex lacteus</i>	<p>NKEFEDVVDV SNLPCGLNCA LYFSSMDADG GMIAKPDGNA GAKYGTGYCDD SOCPRDIKPFI NGEANYAGWN      GSPMDINAGT GNWACCNEM DWEANSTSA AYTHPPTVQ GLSRCSGTAC GNDRYGIVC DPDGCDFNSY      RMGDITYYGP GGTGYDTRK FTVTGYDTLN NISSSGTLESE IRLYVONGR VVONSKVNIP GMNTLSDSTT      TGFCISQKTA FGDRSFQNK GMISAMGQAL GAGMVLVLSV WDHIANMLW LDNHPVDAAD FSKPGLIART      CSTTGSKPTD VEOQAAANSEY TRENKEFEDI GTTYTGGSVT TPGNPGBTIT STAPGAVOTK WGCGGGQWT      GPTRCESGST CTNNQWYSQ CI</p>
SEQ ID NO : 67	50844407 #	Chaetomium var thermophilum var thermophilum	<p>MFRKALLAF SFLIAHGGQ VGTNQAENHP SLPSOHCAS GCTTSSTSVV LDANWRWHT FTGYTNYCTG      QTWDASICPD GVTOKAKACAL DGADYSCSYG ITTSIGNALTI QEVIGTNTVGS RVYLQDASN YQLFKLINOE      PTFDDVMSNL PCGUNGAYYL SQMDQDGAVS PRD1KFINGE ANVAGWTCGS      SDPNSTGNY GTCCESEMIDW EANSVAAYAT PHPCSYNQQT RCTGADCGQD ANTYKGVCDP DSGCDENSFRM      GDQTFIGKGL TVDTSRKFTI VTOFISDGT SGNLAEIRR FYVQDGKVIP NSKVNLAGED AVNSITDIFKC      TQQKTAFGDT NRFDODGKIK OMGAALSGM VIALSLWDDH AANMLWLDSD YPTTADASKP GYARGTCPT      SGVPDVEQS SGSATVTSQN IRKGDLNISTPQ SSSSGPSSS SSSTASQSOFT QPSSGSAQW      GQCGEGYSG ATGCCVSPYC HYVNPYISQC Y</p>
SEQ ID NO : 68	4586347	<i>Trpex lacteus</i>	<p>TETPHPLTWK RCTSGGNCST VNGAVTIDAN WRWHTVSGS TNCYTGNEWD TSICSDGKSC AQTCYVDGAD      YSSTGTTT GDSNLKEFT KHQHGHTNVS YTMVLMENDTK YKMFPLGNE FTPEVDVNL GCGLNGALYF      VSMDDGCM S KYSGNKAGAK YGTGCYDAQI PRDLKFINGE NDANAFGEGRY GSCCSENDIW      EANNRDATAFT PHPTCTIIGOS RCGGNSCGGT YSERTAGVC DPDGCDENAY ROGDKTFGK CNTVDTTKM      TVVTPQFHKN S AGVILSETKPF YVQDGKTLAN AESKTKGPNP NSTIQEWQDA QKTAFGDIDD FNPKGWAQM      SKALEBGPML VMSSWDDHYA NMWMLDSTP IDKAGTPEAE RGAEPITSGV PAELEAQPN SNVLFSNTRF      GPIGSTVPGI DGSTDPSNFTA TVAPPSTTT SVRSSTTQIS PTSDPGCII TQWKGQCGI GYTGCINCA      GTTCTELNPW YSQL</p>
SEQ ID NO : 69	3980202	Phanerochaete chrysosporium	<p>MFHKAFLVAF SLVTTVHGGQ AGTOTAENHP QLSSOKCTAG QCSITSASTSV VLDSNWRVH TTSGYTNCTY      GNTWIASICS DPSCAQNA LDAODYACTG GITTSGDAIT LKFTYQDGSV SRVYMLMEDT NYQMPKLMNQ      EFTFDVDSN LPGCLNGANGAYT FVONDQDGTE SKFPNNKAGA KFTGYCDSO CPD1KFING EANVGNVTT      AGDANSGTSF RGTCQEMDI WEZNSISAY TPHCTVTEQ TRCSGSSDCQ GSDRFNGICD PDGCDENSFR      MGNTFYGKG LTWTISQKFT IVTOFISDGT TAGDNIAEIR YFTYQNGKVY PNSYVQITGI DEVNTSITDF      CTQKTVFQGD TNNEAKGKL KONGEAVKNG MVLAISLWLD S DPTTADPSQ PGVARGTCT      TSGVNSQVEQ QEGSSSVYI NIKFGDINST FTGTLTNPSS PAGPVTTSS SEPSPQSTOPSS QPAQPTOPAG      TAAQPAQCGG MGFPTVCA SPPTCHVLPN YSQL</p>
SEQ ID NO : 70	27125837	<i>Melanocarpus albonotatus</i>	<p>MFMRAALLAF TCLAMVSGQ AGTINTAENHP QLQSQCQTT GGCKPLPLSTKV VLDSNWRVH STSGYTNCTY      GNEVNTSLCP DGKTCAAANCA LDDADYCTSY GITSTGTAIT LKFTYQDGSV SRVYMLMDTT HYOLKLUNQ      EFTFDVDSN LPGCLNGANGAYL LSAMDADGM SKYGNKAGA KYGIGYCDQ CPD1KFING EANVGNVTT      GSNTGTSGYG TCCEMDIWE ANNDAAATP HCTTGTQTR CSGDICARNT GLCDHGDGCD ENFSRMGDKT      FLKGKNTVDT SKPFDVQFL LNTDNTGTVL USEIRIYIQ NGVQIQLNSVA NIPGDPNIS ITDNFCQOK      TARGDNMFA OKGSIKOMGE ALNGMVLAL STWDDIAANM LWLDDSDYPTD KDPBAPGTYAR GNCATTGSP      SDVEQVPN S QVVEENKFG DGAFTSGTS SPNPPGGSTT SPTVTSPTP PPTGPTVQW GCGGGIYSG      STTCAСПYSC Y HVNPPYSC Y</p>

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 71	171696102	<i>Podospora anserina</i>	MYSRATPLIF ASLVLGQGV TYTAERHPSM PIVQVTPAQG CTRETEEVIL DAWRWRHIT IT NGYTNCYTGNN ENNAACPDG ATCKRNCAVD GADYSGTGYI TTPSSGALRL QFVKNDNGQ IVGSRVYIMA SSDKYKLFNL LNKEFTFDV VSKLPGLNG AVYFSEMLED GLKSPGKGN AGAKYGTGYC DSQCPDLIKF INGEANVEGW GGADENSIGTG KYGICCAEMD IWFANSDATA YTPHVSYNE QTRCEGVDCG AGEDRYNTIC DKDGCDTNSY RLGNEFYGP GKTVDTRFP TIVTQFVTD GIDSGMLIKSI HRYVQDGAV VDQINFLESG ECEQKSARG DNNVFGQLLG MPMGNGESLUKK MVLVLSIWDD HAVNNWLDL IEPNDADEQ PQVARGHCDP ADGVATIEA AHPPIAVIIS NIKFGAINST FTAN
SEQ ID NO : 72	3913802	<i>Cochliobolus carbonum</i>	MYRTDAPASL SLYQAARAOQ VGTSTAENDHP KLTWQTCGTT GGTMGSNKSG SVVLDMSMRW AHNVGGYTNIC YTGNWSKTY CPDODSCTKN CAIDGADYSG TYGITISNA LSLIEFVTKGK FSENNGSPTY LMETDTKXOM FNLINKEFTP DVDSKLPGQ LNGALYFEM ADDGGIGKGN NKAGAKYGTG YCDSQCPHD1 KP INGRKANVE GWNPSDADPN GGACKIGACM PENDIWEMANS ISTAYTIPHPC RGVGHQECSD AASCGDGSNR YDGQCDKDG DENSTRMGKV DFYSPGATLID TPKMTVTO FLGSQSSLSE IKRFVYQNGK VYKNSQSAVA GVTGNS-SITES RCTAQKKAEG DTSSPFAALGG LINHGAStarL GHVLINSLWG DHAVNMLWLDP STIPTDADPS KPGAAARTCIP TTSKGPEDVE KNSSDATAVF SNIKRGPIGS TRAQPA
SEQ ID NO : 73	50403723	<i>Trichoderma viride</i>	MYOKALISA FLATAROQA CTIQLAETHPP LTWQKSSGGV TCTQGTSVW IDANWRWHA TNISTNCGD NTWSTLCPD NETAKNCCIL DGAZAYASTVG VFTGKSYAQN VGRARYLWMA DTYQEFPL GNEFRFDVDV SOLFCGLNCA LYTFSMADG GVSKYPINTA GAKYGTGYC SQCPDLKFI NGOANYVGWE PSSNANTG1 GHGSCCSSEM DIMEANSEUSE ALTPHCFCTV GOELCIDGSC GFTYSGDRYGC GTCDPDGCDW NPYRUGNTSF YGPSSFTLID TTKLKLTVTO FETSGVANNY YVQNGTVFOQ PNABLGDSG NSLDDDCFAA EEAEFGCSSE SDKGCLTQPK KATSGMVL MSILDDYAN MILDSTPYI NESTSPDN RSCSCTSSGV PAQLSNSNPN AKVYNSNIKF GPGSTENSS GENPPGENPP GTTCTTRPAP STGSSPQPTQ THYQCGEGIG YSGPTVCASG STCQVLPNPy SQCL
SEQ ID NO : 74	3913798	<i>Aspergillus aculeatus</i>	MVDSPIYKT ALLLSMLATS NAOQVGTYY ETPHSLTWQT CSGSGSCTT SGSVTVIDANW RMVHEVGGY NCYSNTWDS SICSTDITCA SEALEGATY ESTYGVTTSG SSLRNFVTT ASORNIGERL YLLADDSTYE TPKLNRNFEFT EDVIVSNLBC GLNGALIFVS MDADGYSRFE PTNKGAKYK TGYCDSQCPF DLKPIDQGAN TEGMIPSSTD VNACTGNGIS CCEMDIWEA NSISAFFAATV GKTICGNGYS TDTCGCGNYS DITDRCGTC DPDGCDENPY RFGNTNFYGP TVVTQFITHD GKTGTILTEI 1GNGPSTYTA ASGNSTESF CKAKTFLFGD TNFETIGGL MVLVLSLWDD HAADMWLWD SYPTTSCASS PGVARGCTCP TTNQATVYA NYFNSITYVS NIKFTGLNST VSGTSGGSS SSSTTSLTKA STTTSKTT TTTSTSTTS SSSSTVNAQY GOGCCGQNTG PTTCASCTCQNDYTSQCL
SEQ ID NO : 75	66628465	<i>Dictyostelium discoideum</i>	MYRILKSFIL LSIVNMSLQ KIGKLTPVEH PPTMFOKCSE GSCTEIOGE VVYDANWRWV HSAOGONCYT GNTWDPCTCP DDETAENY LDGANEBSVY GTTTSSEDSVR INFVTSQGK NIGSRFLMS NESNYQLEHV LGQETFDVD VSNIUDCGING AILYLVMSDSD GSARAFETNE AGAKYGTGYC DAOPCPDLKIF ISGSANVDGW IPSTINPNTG YGMGSSCAE MULMEANMMA TAVTPHPCDT SSQSYCKSDS CGEAAASNY GGICDPOGD YNPYVMGNTS FFGHNMKIDT NSWIVTVTOF ITDDGSSDGK LTSIIRLYVQ DGNYTISQSVS TIDGVEGNEV NEEFCTNQKK VFGBDSDFTK HGGLAKNGEA LRKGMLVLVLS LWDDYQANML WLDSSYPTTS PTDPGVARG SCPTSGVPs KVEQNPYNAV VVYNSNIKVGP IDSTYK
SEQ ID NO : 76	156060391	<i>Sclerotinia sclerotiorum</i> 1980	MISRYLAISS LIAAARAQQL GTNTAEYHPA LTTIVTIDANW RWLHTTSGYT NCXTGNSWDA TLCPDAYTCA ANCALGDADY SGTYGTTSG NSIULNPFVTK GANTNGSRT YLMAAGSKTO YQLLKLQOE FTFDVIDVSNL PCGUNGLALYF AEMADGEGYS RFTNKGAGQ YTGTGCYDAQC PODIKFTINGQ ANSYGWTSSS NDVNTNTGQY GSCCSMNDIIN EAMKISAAYT PHCSVDQGIG ARYSSLUDAD GCDENSYNG DGFGYGAQLT VDTSTKVFETVY TOFTINDGNT SGTLSEIRRF YVQNGKVIPN SOSKVTGVSG NSITDSRCAA QKTAEGDNTNE PATRGGLATM SKALAKGMYL VMSIWDDISA NMILWIDDISA ASKPSAAGV SRGCSASAGV VPADVEANS P SCPTSGVPs KVEQNPYNAV VVYNSNIKVGP IDSTYK

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 77	116181754	Chaetomium globosum CBS 148-51	GASVYNSNIK WGPININSTYA GRGSNTGSGS GSTTLVSSV PSSTPTSTG VPKYQGCGGS GYTGPTNCIG STCVNGQYY SQCQ
SEQ ID NO : 78	145430535	Aspergillus niger	MYRQATALS FASIVLGGQV GHTAETHPS LPIEVCTAPG SCTKEDTTV LDANWRWTHV TDGYTNCYTG NAMNITACPD GKTGAANCAL DGEYEKTYG ITTPERGAL INFETEENVG SRYULMAGHD KYRFLENULINK EFTMDVDSN LPCLGUNGAVY FSMEDEGM SRFEQNTEA KGTGKYCDSQ DGNSSTGKGY TCCEMIDLWE ANIDATAVY HECKVTEQTR CEDITECGAG GNKEYGPEK TVDTSKPFLP VTCQVTAQDT DIGALOSIRR FYVQDGTVIP INSETVVEGD PTNEITDFC AQQTAFGDN NHFTTIGGP AMKSLERMV LVLSLTDDHA VYMMWLDSDN YPTADPTKPG VARGRCDPAA GPVETVEAAH PDAVIYSNI KI GALNSTFA AA
SEQ ID NO : 79	46241266	Nectria haematoxocca mpVI	MSRQVYRAA LLLSILATAN AQQYGTYYTE THPLSIWQTC TSDGSCCTND GEVYDANWRW WYHSTSSATT CYTGHEWDTI ICITDVTGCA NCZLDGATYE ATYGVYTTSGS ELRLNFTVTOG SSXKNIGRSLY LMSDDSNYEL FKLIGEFTFV DVGSNLPFG INCALYFVAM DADGGPSEYS GNKAKAYKYG GYCDSQCPRD LKPTINGBANC DGWESSNNV NTGVDGHGSC CAEMDVNEAN SISNAFPAH QDSVTSQTMCD PDGCMYNPYR LGMTDFYGFQ LTVDTNSPFT VTQFTTDG TSSGFLTEIK RLTYQNGEVTI ANGASTYSSV NGSSITSAPC ESEFTLFGDR NVDFFKHGLE GMGEAMAKGM VLVLSLWDY AADMWLDSL YPVNWSASSTP GVARGRCSTD SGVATEVAE SPAYAVTSN IKGFGPISTS SGSSTGSGS SSSSSSTTK ATSTTLKTT S TSQSSSSSTS AAQAYGQCGG QGIFTGPTTCV SGYCTYENA YQSCL
SEQ ID NO : 80	1q9h	# Talaromyces emersonii	MYRAJATASA LLATARAQV CMLNTENKPA LTWAKCTSSG CSYVRGSVYV DANWRWAHST SSSTTCYTGQN TWDKTLCPDG KTCADKCCUD GAYFSGTYGV TSGSGNQLNK FVTVGYSTN VGRSLYLMED ENNYQMDLIL GNEFPFDV UNICGGLNA LYTFVSMYKD GKSFRPNTKA QAKYGTGKQ ACQPRDVKI NGVANSBEWK PSDSDRGNAGV GKTYCCPBM DIWEANKLST AYTPHPCKSL TQOSECDAK GGTSYATYA GTCDPDGDFP NPYRQCNKTE YGPSSGENVD ITKVKVTVTO FIKGSDGKLIS EIKRILYQONG KVGNPQSEI ANNPGSSVTD SFCKAOKVAF NDPPDENKEG GWSGMSDALA KPMVLYMSLW HDHZANLMLW DSTYPKGSKT PESARGSCPE DSGDDDTLEK EVPNISGVRS NIKFGPIGST YTGTGGSNPD PEEPEPEEP VGTVPQYQGC GGINYSGPTA CVSPYKCNKI NDFTSQCO
SEQ ID NO : 81	157362170	Polyporus archilarius	EQAGPATAEN HPLITWQET APPGCTTONG AVLDANWRW VHDYNGTYNC YTGTNTWDPTY CPDDETAQON CALDADYEG TYGTSQSSGS LKUNFVTCSN VGSRLYLLQD DISTYQFLKKL NRESFSDVY SNLPCGLINGA LYFVAMDADG GVSXYPNNKA GAKYGTGKCD SCOPRDLKF1 DGEAUNEVGHQ PSENNTANTGI GDHGSSCAAEM DVWEANSISN AVTEHPCDTP QGTMCSGDC GETSYNDRYA GTCDPDGCFD NPYRMGNTSF YCPGKLIIDT KPFITYOTOFL TDDDTDTGTL SEKTRFYLN SNVIPONSD ISGTGTGNSIT TEFTAAKOAA FGDDDDPSOH GGLAMGNAAM QGQNLVMSL WDDYAAQMLW LSDDYPTDAD PTTPGIARGT CPTDSGPVSD VESQSPNSVY TYSNFKGP1 NSTTAS
SEQ ID NO : 82	7804885	Leprosphaeria maculans	MFPTIALVSL SPLAIAYGQO VGHTAETHP KLSVSOCTAG GSCTIVQRSV VLDSNWRWLH DVGGSTNCYT GNTWDLSLCP DPTCANA LIGADYSGT GITTSGNAL LKFTVQGYS SEDDSYTMF NLKNGEFTFD VDMSALPCSH NGALFYTEMID KDGGSQREPT NKAGKKGYTG YCOTOCPHDI KPTINGEANVL DWAGSNDPN AGTHHYGTC NEMDIWEANS MGAATPHVC TVQGQTRCEG DGICDUGCD FNSWNGDQF FLGFGKVTDT SSFTVYUOF ITADNTSGD LSEFRLYVQ NGKVIANEKT QIAGMDAYS ITDDCNAQK TIRBDNTFPE OMIGLATINGD AETGMVLVM SWDHEAKM LWLSDYPTD ADASAPGVSR GPCPTSGDP TDVRSQSPE TWFNSNKTG PIGSTPE
SEQ ID NO : 83			MLSASKAAAI LAFFAHTASA WYVGDQDET HPKLNFQRT GKGRSSCTNV NGEVVIDANW RMLAHRSGYT NCYTSEEMNO SACDNEACT KNGAIEGSDY AGTYGTTSG QNMNLFKPFITK RPTSTNIGAR TYLMKDQSONY EMFOLJNER TEDVDSLQRQ CNGNGLFVUS MPQKGQGAPG AKYGGYCYDA OCARDLKEVR GSANAEGSKTK SASDONSVGK KKGACQAQD WIANSATA LPHSCOPAG YSVCEDTNCG GTYSEDRTAG TCDANGCDFN PFRVGVKDFY GKGKTVDTIK KMVTVTQFVG SGNOSEIKR FYVQDGKVIA NPEPTIPGME WCNTQKTVFQ

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 83	1.21852	<i>Phanerochaete chrysosporium</i>	EEAVDNEFGV GMASMEGMS QGIVLYMSLW DDHYANMLWL DSNMPREADP AKPGVARREDP PTSGGKDPSEV EAANNQAVM FSNKFPGIG STFAHAA
SEQ ID NO : 84	126013214	<i>Penicillium decumbens</i>	MFRATLLAF TMAMVFGOO VGTNTARSHP ALTSQKCTKS GGCSNLNTKI VLDANWRWLH STSGYTNCYT GNOQWATLCP DGKTCANCA LDDADYNTG GITASGSSLSK LOFYTGSNNG SRYTLMAADD HYOMFOUINO EFTFDVDSN LP CGFLNGALY AKYPTNAQDR KSYGTYCDSQ CPDIKFING EANVEGNNAIT SANATGNGYQ TCCTEMDWE ANNDAAAVT P HCTTNAQTR KSYGTYCDSQ CSGSUDTRDT GLCDADGDFE NSFRMGQDF LGKGILVDT KPFVTVTQFI TNGTSAGTL TEIRRLVQN GKVIONSSV 1PGIDPVNSI TDNFCSQRT AFGDTNYFAQ HGGIKOVGBA LFTGMVLIALS IWDYYANML WLDNSYPTNK DPSTPGVARG TCATTSGVPA QIEASPNAY VVFENIKRGD INTTYTGIVS SSTSSSSSS SSTSSSSSS STPTQPTGV TV PQMGOQGG IGYTSTTCA SPYTCVLPN YTSQCY
SEQ ID NO : 85	156048578	<i>Sclerotinia sclerotiorum</i> 1980	MYQRALESA LMAGVSAQV GPKQPETHPP LAWKECTSSG CTSKDGSVVI DAWMRWHSV DGYKNCYTGNN EWDSTICPDD ATCQTNCAVD GAYAGTYGA TTGTDKNSR FYMEDEDKY QNFKLUNKEF TDDVYSLP CGLNGLFVY SMADGMSK VETNKGAKY GTGYCDSQSC RDLKFINGKG NVEGMKESAN DKNAVGPHG SCCEMDWE ANS1STAUTP HPCDTNGQT1 CEGDSCGGTY STRYAGTCD PDGCDNPFR MGNEFSYGPQ KMVDTKSRMT VYQFQTISDG TDGSLKEIK RYVYQNGKVI ANSADVSGL TGSNITSDRG TAQRKTFGDE DVFRTRHGGSL GMGDALGHGM VLUMSLWDD NSNMWLUDGE KYPDAAASK AGVSRGTCST DSGRKSTVES ESGSAKUVPS NIKVGSIGST FSA
SEQ ID NO : 86	156112278	<i>Acremonium thermophilum</i>	MTSKIALASI FAAAQYQIG TYTTEHPSL TWOSCTAKGS CTTQSSTIVL DGMRWTHST TSSTNCYTGNN TWDATLCPDD ATCQNCALD GAYDQGKQI TTSGDSURLN FVTQFANKNV GSRYVLLADN THYKTFENLNN QEPFTDVDS NLPCLNGAV YZTANLADGC ISTSNTAGQ YZTGTGCDSCQ PROCKEINGK ANDGWPSIS NNPNTCVGNV GSCCAEMDILW EANS1STAUTP PHSCDTVTOT VCTGDNCGGT YSTRYAGTC DEDGCDNPY ROGENESFYGP GKTIDTNSVF TIVTOFLTTD CTSSGTLINEI KRFYVQNGKY IPNISESTISG VTGNSNITTPP CTAQTAFGD PTSFSDHGGI ASNIAAFFPAG MVLVSLWDD YYANMLWLDS TYPTKTKGAG GPRGTGCTSS GVPASVEASS PNAVYVSI KYGAINSTFG
SEQ ID NO : 87	21449327	<i>Aspergillus nidulans</i> (also known as <i>Emmericella nidulans</i> )	MYTKAZALAA LIVATVRGQA, CSITAETHPS LQWQKCTAPG SCTIVSQQV ITDANWRWLHQ TNSSTNCYTG NEWDISCIS S DTDATKCCL DGADYDTGYG VTAESNSLNL KFTVQGPYSK NIGSRMYYME SESKYOQPTL LGQEFPTFDV VSMNUGCLNG ALYQVSMDD DSQCPRLKF INQQANIDGW QPSNDANAG LGENGSCCSE MDIWEANIKVS AAYTPIKCTT I GOMTCGDD CGETYSSDRY AGICDPIGCD PNSYTRNGDTS PYGEGKTVDT GSKEPTVTVQF LIGSDGNLSE IKRFVQNGK FCTACTAFTG DTNVEFEERGQ LAGMKAFL PMVLYLSVND DHAYNMLWLD STPTDSTKLP GAARGDQPTI SGVPADVESQ APNSV1YSN TGTPSGNNP GGGTTTTT TTSKPSGPTT TINPSGQPT DVQOCGGQW TGPTVCOQSPY TCKTSNDWY QCL
SEQ ID NO : 88	171683762	<i>Podospora anserina</i> (S mat+)	MYQRALESA LLSYSRASA QTAQEEVHPS LTWORCEASG SCTEVAGSVV LDSNMRWTHS VDGYTNCYTG NEWDAILCPD NESAQNCAV LFVTDYATV ITSNGDSTL KFTVQGPYSK VYTMEDDET YQMFDLINNE FTDFDVSNP PGCHNGALYF TSMDADGELS KEVGTNAGAK YGTGTCDSQ PRD1KFINGL CNVEGMPSD SDANGVGGM GTCPEMDILW EANS1STAUTP PHCDSVTQF MCEGDSGGT YSDDRYGGTC DEDGCDENSY RMGNTRFYGP GAIIDTSRKF TVTQFADG GSLSERKRFY VONGEVIPNS ESNUSGVGN SITSERCTAQ KTAQSFEDIF AQHGLSANG DASAMVLL SIMDDHHSIM MWLDSOYPTD DVVEEHADA SVTSPNLIKEP PIGSTP
SEQ ID NO : 89			MMMKQYQYL AAGS1MTGLV AGQGVGQQT ETBPRITWKR CTGKANCTTVY QAEVVIDSNW RWIHTSGGTN CYDGAHAWNTA ACSTATDCAS KCMEGAGNY QCTYGAISTSG DSLT1KFVTK HEYGTNVNSR FYLMNGASKY QMFELMNNEP TEDVDSLSTVE CGHNSALYFV AMEEDGMRs YPTNKAGAKY GTGYCDAQCA RDLKFEVGGKA NIEGPRESSN DENAGVGPYQ GCAEIDWV SNAHAAFTP HACENNNYHV CERDTCGGTY SEDRFAGGCD ANGCDYNPYR MGNDDFYGRG KTVDTDTKFT VVTRFQDDNL EQFFVQNGK ILAPAPTFDG IPASPNLTPE

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 89	56718412	<i>Thermoascus aurantiacius</i> var <i>levi</i> sporus	FCSTQEDVFT DRNFRREVGD EPQNLAAHLI PMLVMSIWA DHYANMLWLD SVYPPEKEGE PGAAARGPDCAQ DSGVTPSEVKV NYPAKRVWS NTFGPFGTST VNW
SEQ ID NO : 90	15824273	<i>Pseudotrichonympha grassii</i>	MYQRALLFSF FLAARAQQA GTVTAENHPS LTWQOQSSGG SCITQNGKVV IDANWRWHT TSGYTNCTG NTWDSICPD DVTAQNCAL DGADYSCTG VTTSIGNALR NEVTOSSGN INGRSLYLIQD DITYQIFKLQ QGEFFFDVVD SNLFCGLNQA LIVFAMADG NLSKYFGNKA GAKYGTGTYD SQCPRDLFPI NGQANVEGMQ PSANIPNAGV GNHOSCCAEW DWEANSIST AVTPHPCDTP QOTMCGDDC GGYSSTSYA GTCDPGCDP F PYROGNHSP YGPCKIVDTIS SKETVUVQFI TDDGTFSGTL TEIKRFYVQN GKVIPOSEST ISGVTGNSIT TEYCAQKAA FGDNITGFETH GGJOKISQAL AGQMVLYMSL WDDFHANMLW LDSTYPTDAD PDTPGVART CPTTGVPAV VESOPNPNTV IVSNIKVGPI NSTFTAN
SEQ ID NO : 91	115390801	<i>Aspergillus terreus</i> NIH624	MFATVLLGLT RSLGTTGMA ENPLSLSWN CRSGGSQT SGSTVYLDSSNN RWTHDSLTLN CYDGNEWSSS LCPDPKTCSD NCILGDADYS GPTGTTSSGN SIKLVFVTNG PYSTINIGSRV YLKJDESHQ IEDLKNKEFT FTVDNSNLC GLNQALYEVs MDBDGGTSRF TGYCDAQCPH CPTDQCDSO DGSECQDTS GNGCDENSYRM ENAGNGRYGA CCTPEMDINEA NKYATAATPH ICTVNGEYRC DNGQLSIRR KIVQGKVKIE INTVNLAGMS SENSIITDFC NEQKAFGDT NDPEKKGGIS GLCKAFADYGIV VLVLSWDDH QVNMLWLDI YPTDQASOP GYKRGPCTAS SGAPSDEVSQ HPDSSVTFSD TREGPIDSTY
SEQ ID NO : 92	453223	<i>Phanerochaete chrysosporium</i>	MHORAILLFA LVGAVRAQQA GTLUTEEVHPP LTWQKCTADG SCTEQSISV IDENWRWLHS TNGSTNCTG EFTFDYDVSN LPQGUNGALY FTSDGTSRFP TSNDGTSRFP SKYPAKNGA KYGTEYCDSO CPTDQCDTS GLCDCGCDPNA DNDRAVQGG HGSCCPEDI WEWNSLISAP TPHCDLQG TYSETRFAGT CDPPGCDPNA YRMNTSYGG PDKVDTNSV MTTVTFQIGD GSLSLEKRL YVQNSKRVIAN AQENVDGTV NSITSDRQTA QKTARGDQI FSKVGGLSM GDZMSAMYL LSIWDDHNSS MMWLDSTYPE DADASEPGYA RGTCHEHGVD PETVSOHQPG ATVTFSKIRK GPGSTYSSN STA
SEQ ID NO : 93	3132	<i>Phanerochaete chrysosporium</i>	MFRAAALLAF TCLAMVSGQ AGINTAENHP QLOSOQCTTS GGCKPLSTKV VLDSNWRVWH STSGYTNCTG GNEWDTSLCP DGKTCANCA LDGADYSGTY GITSTGTAII LKPTGNSV SRVYLMADDT HYQLKLINQ EFTFDYDMSEN LPQGUNGALY TSNDGTSRFP TSNDGTSRFP SKYPAKNGA KYGTEYCDSO CPTDQCDTS GLCDCGCDPNA ISNDAAAPF HCTTGTQTR SGDDCART LGKGTIVDTIS KPFVTVTOFL TNDNTSTCL SEIRRLYTON GVTONSVAN IPGYDPVNSI TDNFCAQOKT AFGDDINWPAQ KGKIKQMGEA LGNGMVLALS IWDHHAANML WLDSDYPTDK DPSAPGVARG TCATTSGVP DVECSYPNNSO VVFNTIKFGD IGTEFSGTTS PNPGGSGTTS SPVTSPTPP PTGPTVPOKG QCCGIGTSGS TTCAFSYTC VLNPCEISLQ LESSNADQY LQTRSATKR RLDTALQPRK
SEQ ID NO : 94	16304152	<i>Thermoascus aurantiacius</i>	MRTAALALIA LAAPSAVQO QAGTITAETH PTLTIQCTQ SGCCAPLTIK VLDVNRWMI HSTTGTYNCY SGNTDAILC PDPTCAANC ALUDADYGT EGILPCTSV TLRPVTDGLL RLFLDADISH YOMFOLINKE PTFDPEMPNN RCGSGAHLA TAMDADGELA KYPGNQAGAK PKGKFINGQ ANVEGMGFT ATTGGTGFRRGS CCTTIALHBA NDUNASAFPH PCTTNEOTRC SGSDCTADSG LCDADGMENI SFRMGNTTFP GAGMSVDTIK LFTVVTQFIIIT SDTISMALV EIRLIVTQNG QVIONSVVN PGINPATSTT DDLAQENA EGGTSSPAOH GHIAQVGEAL RSGMVLALSI VNSAADTML DSNTPADADP SAPEVARGETC PODSASPEA PTPSTVFSNI KLGDIGTGF AGSALFSGRS PPGPVDSAP ASSATAPP FGSGCQGILGY AGPTGVCPSP YTCAQALIYV SQCI
SEQ ID NO : 95			MYQRALLSF FLAARAHA GTVTAENHPS LTWQOQSSGG SCITQNGKVV IDANWRWHT TSGYTNCTG NTWDSICPD DVTAQNCAL DGADYSCTG VTTSIGNALR NEVTOSSGN INGRSLYLIQD DITYQIFKLQ QGEFFFDVVD SNLFCGLNQA LIVFAMADG NLSKYFGNKA GAKYGTGTYD SQCPRDLFPI NGQANVEGMQ PSANIPNAGV GNHOSCCAEW DWEANSIST AVTPHPCDTP QOTMCGDDC GGYSSTSYA GTCDPGCDP F PYQOCGNHSP YGPCKIVDTIS SKETVUVQFI TDDGTFSGTL TEIKRFYVQN GKVIPOSEST ISGVTGNSIT

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 95	156712280	<i>Acremonium thermophilum</i>	TEYCRAOKAA FDNTGFFFTG GLOKISOALA QCMLVLYNSLW DDHAANMLWL DSTYPTDADP DTPGVARGTC PTTSGPADY ESDQPNSTVI YSNIKRGPIN STFTAN
SEQ ID NO : 96	5231154	<i>Volvariella volvacea</i>	MHKRATLSEA LIVVAAAGFAF GCGVGTTQOEE THPLKTPQKC SAAGGCTTQN GEVIDANWR WYHDKNGYT CYTGEMWNNT ICAIAASCS NCVDGADYO GTYGASTSGN ALTLKFVTKG SYATNIGERM YLMAASPKYA MFTLIGHEFA FDVDSLKPCF MDGGTSPCR TGCDQSCLR DLKFIDGKAN SASWPSND QNAQVGKMGCS CCEZMDIWEA NSVSAAYTPH PCQNYQOHSC SGDDCGETYS ATREFAGICD P DGDNAYRM GVHDFYGNK TWTFGKRFSTI VTFQKSGST LTETIKOFYVQ DGKJENPA TWGLEPENS ITDPDKAOK QVFEDPDREN DNGCFTNMKA ALANPMYLVL SLWDDHYSNM LWLDSTYPTD ADPSAPSKGKRG TDGDSSSGPV SDVSESKNGDA TVIYNSIKPG PLDSTYTA S
SEQ ID NO : 97	116200349	<i>Chaetomium globosum</i> CBS 148-51	MRASLIAFLS NSAGQQQAGT LQTKNHBPBLT SQKCRQGCP QVNTIVLDA INWRTHSTSG STNCYTGRNW QATLGPDGKT CAANCALDGA DTGTYGWT SIGNSLILQFY TQSNGARLG YLMADDTYYQ MNLLNQEFW FDVDSNLPC GLNGALYSA MARTAAMPV VCASTPLS TRSTARLRL LPVPPRSRSG RGCIDSQCP RYGSICDHDG LGFRNLFENG RTVRAVGR VRQFNRSRV VEPISTWTKQT TLHUGNLWK SADCNVQRGR VIQNSKVNP GMPTSMDSVT TEFCNAQKTA FNDTFSRQK GMAMNSEAL LDSTSAAAC RSTSEVERAT PUFESQRSS HSRQTYVFT TNIRKGFPNS TGTTYTTGSV PTSTSTGTG TSSTSPQPTG VTVFGQOCGG IGYTGPCTCA SPITCHVLPN YSYCQ
SEQ ID NO : 98	4586343	<i>Trpex lacteus</i>	MKOYQYLAAL ALPLMSLYSA OGWTGSTSET HPKITWKKCS SGGSCSTVNA EVYDANWRW LHADSKNCY DGNEWTACT SSDDOTSKCV LERAEYGTG GASTISCDLSL IKFLITKHEYG TNGSCARPLK NGASKYONFT LMNNPFAFDV DLSITVECCGCL DSGMASYNT KAGAKYGTG CDQSCARPLK FGCKANVDC WTPSNDANA GVGAGGCCA EIUDWEENAH AFAFTPHACE MNNTVHCVDT TCGETYSED R FAGDCDANGC DYNPRVENT DEYGKEMTD TSKKFTVYSO RQENKLTOFF VONGKKEIP GPKHEGLTE SSDITPELC S AMPEVFGDRD RFAVGGFEDA LNKALATPMV LMSIMDHY ANMLWLDSSY PPERAGTPGG DRGPCAQDSG AVPSYEQY DATVWSNIR FEGIGSTVQ
SEQ ID NO : 99	15321718	<i>Lentinula edodes</i>	MFPKSLIAL SFLIAVYGGQ VGTOMAETHP KLPQLCCKS GCTHQNTAVY LDANWRWHT TSGYTCYTG NSWDATLCPD ATTCAQNCAY DGDYSGTYG ITTSGENALTL KFKTNTVNGS RVYLMQDFTA YOMFOLLINEQ FTFDYDMSNL PCGILINGALYL SQNDGSLS KEPTRNAGAK DCGMASYNT KAGAKYGTG GDGSQGDD DERYSGLCKD DGCDFENFRM GDKSFHICKGM TVDLSRKFTV VTFQVFTDGT TNGDLHEIRR LYVQDGKVIQ NSVSIPIGD AVDSITDNC AQOKSYFGDT NYFTLGGK KNGQALISGM VLAMSYWDH ASHOMLDSN YPADGDTK P GYARGTGSAD SGLPNIVESQ SASASVTSN IKWQDINTF TTGTGSTSPSS PAGPVSSSTS VASOPTQPAQ GTVAQWQCG GTGFITGPTVC ASPPTCHVNPYTSQY
SEQ ID NO : 100	146124875	<i>Pleurotus</i> sp Florida	MFRTAALLSP AYLAIVYGGQ AGTSTAETHP PLTWEQCTSG GSCTTOSSV VLDNSMRWTH VYGGYTNCYT GNEMNTVCP DGTCAANCA LIGADBEYQ GISTSGNATL KFKTNTVNGS RVYLMQDFTA YOMFOLLINEQ NPLNQEFTEFV DVSALPCSH NGALYFSEMD ADDGLSEYPT NKAGAKYGTG YCDSOCRPDI KPIEGRANVE GWTPTSTSPN AGTGTGGLCC NEMDIWEANS ISEALTPHPC TAQGFTACTG DCSGPNSTA GICDQAGCD P NSFRGDTSP YGPQHITVDT SKLTVVQFI TSDNTTGTDL TAIIRLYVN GOVLONSMN TAGVTPTNEI TTDFCDQKT ARGDINTSEB KGGLTGNGAA FSRGMVLVS WDDIAAEML WLDSSTPYGR TGPGAARGTC ATTSSQPODV ETQSPNAQVY RENIKRGAIG STESSSTGQV GTGIGSTGT GTTSSSAAA TQTKYQCGG QGWIGATVCA SGSTCTSSP YYSQL
SEQ ID NO : 101			MFRTAALLAF TFAVVLGQG VGTLLTENHP ALSIQOCTAT GCTTQOQSIV LDNSMRWTHS TAGATNCYTG NAMDDALCPD PATATNCALI DGDYSGTYG ITTSGENALTL REVTGQYQSY NIGSRVYLLD DADHYKLFDL KNOETEFDVD MSGPQCGLNG ALYFSEMAAD GCKAAHAGNN AGAKYGTGCV DAQCPHDITKW INGEANLYDW SASATDDNAG NGRGACCAE MDIWEANSEA TAYTPHVCRD EGLYRCSGTE CGDGNNRYYGG VCDKDGCDFN

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 101	62006158	<i>Fusarium venenatum</i>	<p>SYRMEDKINFL GRKTIDITK KVTVTOFIT DNTPTGMLV EIRRYVONG VVYONSFSTF PSLQXNSIS S DEFCAQKTL FGDNQYHNT GGTYKMDA DNGMQLMSL WSDHAHMLW LDSDYPLDKS PSEPGVSRGA CPTSGDPPD VVANHPNASY TRNIKYGPI GSTFGSSTPP VSSGSSSVPP VTSTTSSGTI TPTGPTCTVP KWQGGIGY SGPAVAGVS TCTYSNDWYS QCL</p> <p>MYRALTASA LIAVRAAQV CSLTPETKPA LSWSKCTSSG CSVNQGSVTI DAMWRTHOL SGSTNCYTGN KWDTSCTSG KVCEKCCUD GARYASTYGI TSQSGNLSLS FVTKTYGTN IGERTYLMED ENTYOMIQL GNEFTFDVDV SNICGLANGA LYFVSMADG GRAKYPENKA GAKYTGYCD ACPRDVKFI NGQANSIDMQ PSKSYNGGI GNLCCTCCPEM DWEANSTST AHTPHPTCKL TOHSTGDS SC GGTYSEDEYG GTCDAQGDFP NAYQENKTF YGPSSGRNVD TPKVTVTQ PHKGNSNRLS ETTRLYVNG KVJANESKTI ACVPGSSTLP EFCTQKVKV GDIDDEFEKK ANGMSDALK APMLVYMSLW HDHHSNMLW DSTYPTDSTK LGAQRGSCST SSGVADLEK NVPSKVAEFS NKFGQPAPN PTNPAPPTPG GTVDQWGCQK GNYSGPTAC KSPFPICKKIN DFYSGCQ</p>
SEQ ID NO : 102	296027	<i>Phanerochaete chrysosporium</i>	<p>MERTATLLAP TMAAMVEQQ VGTNTAAENHR TLTSQKCTKS GGCSNLNTKI VLDANWRMLH STSGYTCY CTGNTQATLCP DGKTCANA GA LDGADYDTY GITASGSLK LQFTGGSNVG SRYLMLADTT HYQMFQILNQ EPTFDVDMSN LPCGUNGALY LSAMDAGGM AKYPTNKAGA KYGTGYCDQS CPDIKIFING EANEGHTNAT SANAGTGNYG TCCFMENDIWE ANDDAAATT PFCITINAQTR CSGSDOTRDT GLCDADGDFN NSFRMGDOOTF LGKGLTVDTS KPFVFTVQPI TNDGTSAGTL TPCIDLNSTI TDNECSQQT AFGDNTYFAQ HGGKROVGEA LRTGMVLALS IWDYANML WLDNSYPTNK DPSTPGVARG TCATTISGPA QIEAESPNAV VPERNIKEGD LNTTYGTIVS SSVSSSSHHH TSTSSSSSS STEPPTQPGV TVPQWGCQG IGYTGSTTCA SPYTCVHLNP YTSQCY</p> <p>MYOTSLLASL SFLIATSQAO QVGTQTAETH PKLTOKCTT AGGTQDQSTS IVLDANWML HTVDGYTCY TGQENDTSC TDKTCAERK ALLGADTEST YCISSTSGNAL TMNFVTKSSQ TNIGRIVYL AADSDDTYEL FKLKNQEFPT DVDSNLPGC LNCALYSEM DSQGATGSKYV TINKAGAKYGT GYCDTOCPHD IKPLINGRANV QNWTSSTDK NAGTCHYGS CNEMDIWEAN SQATAFTPHV CEAKEYQVR CEGTECGDQD NYGGV/CDKD GCDFTSYRMG NETYGSNS TLTDTKKEPTV VTOFIFTADT ATGALETEIRR KYONDVYIE NSYADYTLS KFNS1TDDFC AAQTKLSEGT NDFTKGGIA RMGESFRGM VLVMWSVDDH AANALWLDSS YPTDADASKL GVKRSPCSTS SGVPSDVEAN DABSVIYSN IRYGDGSTF NTKA</p>
SEQ ID NO : 103	154449709	<i>Fusicoccum sp</i> BCC4124	<p>MYOTSLLASL SFLIATSQAO QVGTQTAETH PKLTOKCTT AGGTQDQSTS IVLDANWML HTVDGYTCY TGQENDTSC TDKTCAERK ALLGADTEST YCISSTSGNAL TMNFVTKSSQ TNIGRIVYL AADSDDTYEL FKLKNQEFPT DVDSNLPGC LNCALYSEM DSQGATGSKYV TINKAGAKYGT GYCDTOCPHD IKPLINGRANV QNWTSSTDK NAGTCHYGS CNEMDIWEAN SQATAFTPHV CEAKEYQVR CEGTECGDQD NYGGV/CDKD GCDFTSYRMG NETYGSNS TLTDTKKEPTV VTOFIFTADT ATGALETEIRR KYONDVYIE NSYADYTLS KFNS1TDDFC AAQTKLSEGT NDFTKGGIA RMGESFRGM VLVMWSVDDH AANALWLDSS YPTDADASKL GVKRSPCSTS SGVPSDVEAN DABSVIYSN IRYGDGSTF NTKA</p>
SEQ ID NO : 104	169359460	<i>Coprinopsis cinerea</i> Okayama	<p>MFSKVALTAL CELAQAQAO VGREVAENHP RLFWQRCTR GGCTVNSQO VVLDANWBWL HYTDGYTCY TGNAMNSVC SDGATCAERC ALLEGANTQQT YCITTSQDAL TIKFLTRSEQ TNICARVLM ENEDRYQMNEN LLNKFETEDV DVSVTPCGIN GALYFIQMDA DCGLSSOPNN RAGAKYGTGY CDSOCPRDJK FINGEAMSVG WEPSETDPNA GKGQYGICCA EMPIWEANI SNAVTPHPCQ TNDGTYQRC QGRDVNPBY EGICDPDGCD YNPFMNGKD FYGKVTVDI NERMTVYTOF ITHDNTDTGT LVDIIRLYVO DGRVIANPPT NPGGLMPAHD SITQFCCDDA KRAPBDNSF GRNGGLAHLNG RSLAKGHVLA LISWINDHTAH MLWLDNSYPT DADPNKPGIA RGTCPTGGS PRDTEQHFD AQVFSNPKF GDIGSTFFSGN</p>
SEQ ID NO : 105	50400675	<i>Trichoderma harzianum</i> (Anamorph of <i>Hypocrea lixii</i> )	<p>MYRKLAIVSA FLAARAQOY CPOOAEATHPP LTWOKCTASG CPOQGSVVL DANWRWHDY KSTTNCYDGN TWSSTCPDD ATCAKNCUCL GANYSGYGV TISGDAITLQ FVTAENVGSR LYIMANDSTY QFTLSCNEP NFTDVSOLP CGUNGALYEV SMADGGSOK YEGNAAKY GTGYCDSOCP RDJKFINGSA NVEGMWESSN NANTGVGHG SCSEBMIDWE ANSISEAUTP HCETVQOTM CGDSCGGTY SNEYGGTCD PGCDWMPYR LGNTFVYGPQ SSFALDTPKL LTIVTOPATD GSLSRYYVON GVKEQPNQAO VGSYSGENTLN TDYCAAROTA FGGTSTDKG GLAQJNKAQF GENVLYMSLW DDAYVNMLW DSTYPTNATA SPEAKRSSC STSSGVAQV EAQSNNSKVY YSNURFQGPIG STGGNTGSNP PGTSISTRAPP SGTGSSPTAT QTHYQGCGST GMGTPTRCAS GYTCQVLNPF YSQLN</p>

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 106	729649	Neurospora crassa (OR74A)	MRAISLIAFLS AIAVAGGQQA GHLIAKRRPS LIWIKCITRGG CPTLNITMVL DAWRWTIATAT SESTKCYTGN KWQATLCPDG KSCBANCALD GADYTGTYGI TGSWSUTLIO FVTDNYGARA YLGADDTTOYO MLELLNGBLW FDVDSNIPC GLNGALYUSA MDDGGMKY PINKAGAKYA TYCDAQCPR DLKTYINGTAO VEGWTPSTND ANGICBDHGSC CSENDIWEAN KVSTAFTPHP CTIIECHMCE GDSCCGTYSD DRGVLCDAAD GCFDNSTRMG NTTFGEGTK VDTFSKFTVV TQFISKSQD LAEIKAFYVO NVKTIENQS TOSFKSQKT TOSFKSQKT AFGDDDENK KGKHKOMCRA LAGAMULAMS IWDKHANML WLDSTYVPKV VPGLAYRGSGP TTSGVPPEVD ANAPSKVAF SNIRKGHLGI SPSSGGSSGT PPNSPSSAS PTSSTAKPSS TSTAQNPSGT GAHHWAQCGG IGFSOPTTCP EPYCIAKHDH IYSQCV
SEQ ID NO : 107	119472134	Neosartorya fischeri NRRL 181	MLASTFSYRM YKTAALILAA LGSQGAQQVG TSOAEVHPSM TWOSETAGGS CTTNGKVVI DANWRWYHVK GDYTCYTGNN TWDTKLCPDD ATCASNCLAE GANYQSTYGA TTSGSLSLRN FVTISQKNI GSRLYMMKDD TTYENFKLNN QEFEDVDS NLFQGLNGAL YFVANDADGG MISKYPPTNKAG AKYTGTYGCDS QCPRLDKFEN GQANVEGMQP SSNDANAGTG NHSCCAEMD IWEANSISTA FTHPCDTPG QWMCYTGDAAG GRYSSDEYGG TCDPGCDEN SFROGKDNF KETVYQTQIT DGTGASGTLK EIRFYYVQNG KVLPNISESTW SGVGGNSITN DYCAQAKSLL KDNQVFAKHG GNEGMGRALA QGMYLVMSLW DDHANMMLW DSNPITTAAS STPGTARGTC DISSGVPADY EAHMPDASVU YSNIKYGPIG STENSGGSNP GGGTTTTAKP TTTTTAGSP GGTGAQHYQ QCGCNGWQGP TTCAСПYTCQ KLDNFYSQCL
SEQ ID NO : 108	117935080	Chaetomium thermophilum	MQIKOLOYL AAALPLVNA AAORGATQOT ETIHPRLSWKR CSGGENCOTV NAEIVIDANW RMLHDSNYQN CYDGRBWTSA C5SATDCAQK CYLEGANYG TYGVTSISGDA LTLLKEVTKHE YGTMNSRSVY LMNGSDIQYQ FTLMNNEFAF DVDISKVECG LNSKALYFWM BEDGGRNFSYS SNKAGAKYGT TYCDAQCARD LKFVFGGRANI ECWRSITNDA NAGCPCYGC CADIWESN AYAFATPHLQ CLNNHYVCE TSNCQGTYSE DRFGCLCDAN GCDYNPYRMG NKDYKGKGT PHSCDTVQT VDTSRKFTVV TFEENRHLQO FFIQGRKID IPPPTWPGHP NSAITPELC TNLSKVEDDR DRYETGGER TNEALRIPM VLVMSIMWDGH YASMLMLDSV YPBEKAQGPG AERGP CAPTS GVPAVEAQF PNAQTVIENI RFPPIGSTYQ V
SEQ ID NO : 109	154300584	Botryotinia fuckeliana B05-10	MTSIALVSL FAAYGGOVG TYCOTETHPSL TWOSETAKGS CTTNTGSIVL DGMRWRWTHGV GTSTNCYTGNN TWDATLCPDD ATCQNCALAE GALNSGTYGI TTSGSLSLRN FVTSQANKNI GSVYLMADT THYKTFENLN QEFDPDVDS NLPGCNGAV YFANLPADGG ISSTNTAGAE YGTGTYCDSOC PRDMKFKIQO ANVDGWPSS MNANTGVGNH GSCCAEMDLW EAMISITAVI PHSCDTVQT VSSSRVAGTC DEPGCDENSY RMDGDTFYG P GKTWDTNSVF VTVTQFLTGF DTGASGTNEI KRFYQDGKV IPNYSITSG VSGNSITTP CDAQKTAFGD PTSESDHGGI ASMSAEEAG MVLVLSIMWDH YANMLMLDS TYPGKTSAG GPRGTCDTIS GVPAVEAASS PNAVYVSNI KYGAINTSYG
SEQ ID NO : 110	15824271	Pseudotrichonympha grassii	MFVFVILMLT OSLIGTGNOA ENHPSLWN CRSGGSCTOT SGSYVLDSDN RWTHDSLNTN CYDGENESS LCPDPKTCSD NCLDGDAYS GTYGITSGN SLKLKFVTING PYSTINISRSV YLKKDESHYQ IFDLKNEKF FTVDSNLDLDC GLNGALYFVS MDRDGGTSRF SSNKAGAKYGT TYCDAQCPR DLTIFNGRAN VENWKPOTND ENAGIGRYGA CCTENDIWEA NKVATAYTPH ICCTVNGYRC DGSECGDTS GARYGGVCDP DGDFNSYRM GNTSTWGPGL IIDTGPVPTV VTOQFTKDGT DNGOSEIRR KYVOGKVKIE NTVUNIAGMS SENSITUDFC NEQKSAPGDT NDFERKGGLUS GLCKAFDGYM VLVLSEMDDH QVANMLMLDSI YPTDQPSAOV GVKGPCATS SGAPPDVEQS Q HPDPSVTFED IRGPIDSY
SEQ ID NO : 111	4586345	Lepid lacteus	MERKALLAR SPLIAJAHGQO VGTNQAEHNP SUPSOQKOTAS GCTTSSTSIV LDANWRWYHT TTGTYTNCYTG QTWDASI CPD GVTAKACAL DGADYSGTYG ITTSQHNLTL QFVKTGTVNGS RVYLLQDASN YOMFOLINEQ FTEDYDMNSLU PGCHNGAYL SONDQDGVS REPTNTAGAK YGTGTYCDSOC PRDIKETINGE ANVEGNTGS TDSNSCTGNY GTCGSEMIDLW EAMSVAAATY PHPSCTYQNT RCTGADGQG DGCFDNSPRM GDQTEFLGKL TVDOSRKFTI VTOQFTKDGT TSGNLAEIRR FVYQDGQNVIP NSKVSIAQID AVNSITDDFC TQOKTAFGDT NRFAQGGK QMGAALKSGM VLALSIMWDH AANMLMLDS YPTTADASNP GVKGPCATS

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 112	46241268	Gibberella avenacea	SGFPDVEQ SGSATVTVSN IKWGDLNSTF TGTLLTPSGS SSPSPASTS GSSTSASSSA SYPTQSCTVA QWAQQGIGY SGATCVSPY TCHVNNAIYS QCY
SEQ ID NO : 113	6164684	Aspergillus niger	MYRALTASA LIAARAAQV CTTTEKPA LIWSKCTSSG CTDVKGSGI DAWRWTHOT SSSTNCYGN KWDTCYCTSG ETCOKCCUD GADYAGTYGI TSGNOLSLG FVTGSESTN IGBTYLMEN ENTYOMPOLL GNEFTFDVDV SNICGGLNGA LIVFSDMAD AQCRDVFIFI NGKANSIDWK PSDSINAGI GMOTCCPBM DIWEANSTUT AETPHPTKLA TQAHATGDS NSYQRNKTPE YGRSDENDV TPKVTVTQ FRKGNSNRLS ETIRUYQONG KVIANSESKI PGNSGSLLTA DFCSQKSVP GDIDIFSKKG GMGMSDALE SPMVLYMLW LDSTYPTDST KLGAGRSSCA TISGPDSL RDVNSKVSF SNTKPGPIGS TYSSTGTNPP PSSTDISTTP TNPPGGTVG QYGQCGQTY TGPCKSKPY TCKRKINDFYS QQQ
SEQ ID NO : 114	6164682	Aspergillus niger	MSSEPIYRAA LLLSILATAN AQQGTYTTE THPLSLWTC TSDGSCTTND GEVYUDAWR WHTSTSSATN CYTGHEWDT SICTDVTCZA NCALDGAIVE ATYGVTSYLS ELRNLNEFTQG SSNTNGSRAY LMDDDSNTEL FKLICQEFET DVDSNLPCG LNGALFYAM DADGGESEYS GYCDSOCPRD LKFRINGEANC DGWEESNNV NTGQDHGSC CAMPDVHNAN SISNAFTAHP CDVSQTMCD GDSSCGTYS A SCDRYSECTCD PDGCDYNPYR LGNTDFYGPQ LTVDTNSPRT VVTOFIIDG TSSGLTEIK RLYVQNGEVII ANGASTTSSV NGSSTSACF ESEETLFGDE NYEDKHGLE VVFLWLDDY AADNLWLD SD YPVNNSASSTP GVARTGSTD SEGVATVEAE SPNAYVTSN IKFEGPIGSTV SSSSSSTTK ATSTTLKTS TTSGGSSSTS AAQAYGQCGG QWTGQPTTCV SGYTCTYENA YSQCL
SEQ ID NO : 115	33733371	Chrysosporium lucknowense	MHQRAILFSA LLTAVRAAQV GTLVEEVHPS LTWQKCTSEG SCTEQGSVV IDSNWRWTH VNDSTNCYTG NTWDATLCPD DEPAANCAL DQDYEVTG VITDGDLSL KEVPOISNGV RLIIMDTSDE GYOTFNILLDA EFTFDYDVSN LPGCUNGALY FTIAMDADGCV SKYPANRAGA KYGTEYCDSO CPRLKFIDG QANVGDWEP SNNNDTGIGN HGSCCPENDI WEZANKISTAL TPHCDCSSEQ TMCEGNDCGG TYSDDRYGGT YRMGNDSFYQ PGKTIIDTGSK MTVUTQFIDT GSGLSEIJKR YYVQONGNVIA NADNSIISGYT GNSI TTDFCT AQKKEFGDED IFAFHNGLAG ISAMMSMVL ILLSWDDYYA SMEWLSDYIP ENATADPVG ARGTCDSSESG VPATEGAHP DSSTYFNSIK FGJNSTSA SA
SEQ ID NO : 116	29160311	Thielavia australiensis	MYAKPATLAA LVAGAAQAQV CTTAENHPS LTWQKCTSGG SCTSYQGSIT IDANWRWTHR TDSATNCYEG NKWDDSYCSD GPSASKCCI DGDYSSTYG ITTSGNSLNL KEVTKQYQST NIGSRTYLME SDTKYOMFQL LGNEFTFDVD VSNIQGGLNG ALIVFVSMDAD GGMSKYSGNK AGAKYGTGYC DSOPRDLKF INGEANEVNTW QSSTDANAG TGKYSGCCBE MDWWEANIMA AFTPHPCXV IGQSRCBEGDS CGGYSTDRY AGICDPGCD ENSYQGNKT FYHGGMIVDT TKITVYUOFL LKNSAGELSE IKRFVYQONG VPINSESTIP GVEGNSTOD WCDRKAAPG DVTIXQDKKG MWONGKALAG PWLVMSIWD DHAYNMWLW STWDGACK PQAERGACPT TSGVAAEVEA EAPHSVNFPS NURFGP1GST VSLGPUCGSG NPNPVEVSSST PVSSSSTTSS GSSGPTGOTG VAKHFEQCGG IGFPTQCB SPVCTKLND WYSQCL
SEQ ID NO : 117	146197087	uncultured symbiotic protist of	MYAKPATLAA LVAGAAQAQV CTTAENHPS LTWQKCTAPG SCTVAGS IT IDANWRWTHQ TSATNCYEG SKWDDSYCCTT GTDQASKCCI DGAEYSSTYG ITTSGNSLNL KEVTKQYQST NIGSRTYLME SDTKYOMFQL LGNEFTFDVD VSNIQGGLNG ALIVFVSMDAD GGMSKYSGNK AGAKYGTGYC DAQCPDLKF INGEANEVNTW ESSTDANAG SGKYSGCCBE MDWWEANIMA TAFTPHPCCTT IGOPTCEGDT CGENYSSDRY AGYCDPQGCD ENSYQGNKT FYHGGMIVDT TKITVYUOFL LKNSAGELSE IKRFVQDGL DHAYNMWLW STWPDOQGV AGAERGACPT YCDAGTKVRO NTDDEFTAKGG LVQMGKALAG DMVLYNSVMD DHAYNMWLW STWDGACK PQAERGACPT TSGVSDVEA NAPHSVNFPS NURFGP1GST VQGLPSSGT SSSAAQPS TSVKASITTS AVRTSTATT KTTSSAPAQG TNTAKHWQOC GGNGWTQPTV CESPYCCTKQ NDWYQCL
SEQ ID NO : 118			MLTIAVFLLS LIVSYLEIGTO QSDDHPKLW QNGSSVSGS IVLDSNNWRWV HDSGTTNCYD GNJWSKDLCP SSDTCSQKCY TEGDADSYR GROSSGSLT LKFVTRGSYS TNIGSRVYL KDENTYESPK LRNKEFPTTV

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
		Reticulitermes <i>speratus</i>	DDSKLNGGLN GALFYVAMDA DGGAKYKSSF KPGAKYGMGY CDAQCPHDMK FISGKANVDD WKPQDNDENS GNGKGTCGS EMDIWEGRM SQAYTVHACS KSGQYBCTGQ QCGDFTSGDR FKGTCDFDGK DYASWRGQDQ SFGYGRKTVD TKQPVTVYTO FICGDPLTEIR RLYVOGKTIK NMSKTSNLAD TYDSITDFKC DATKEASGDT NDFKAKGAMS GFSTNLNNGQ VLMWSLWDDH TANMLWLSDST YPTDSSDSTA QRGPCTPSSG VPKDVEQSQHG DATVYSDIK FGAINSTFKW N
SEQ ID NO : 118	146197237	uncultured symbiotic protist of <i>Neotermitess koshunensis</i>	MLAAALFTRA CSYGVGTTKTP ENHPKLNWON CASKGCSQSV SGEYTMDSNW RWTHDENGKNI CYDGNTWJS LCPDPDKTSD KVLDGAAETQ ATYGQISNGT ALTLKVTHG SYSTINISRLN YLLKDKSTYY VPKLNKEFPT FSVDSKLPC GLNCALYFEV MDADGGTAKY AGAKPAGAEY LGYCAQCPVS DLKFVTKGSYS TNIGSRVYL KNAGRGKYG CSMSDWNVES NSCATLPH VCKTTCQRCR SGKSECGQGDN EDGCDFNINWR MGDKTFFGPG LIVTKSPFV VYTOFYSPV TEIRRKYVQN GKVLLENKSN IPGIDATAAI SDHFCEEQKMK AFGDNDFKN KGGFAKLPGV FDGGMVLVLS LMDDHOVAML WLDSYTPNKK DKSGPQYDQRG PCPTSSGKPD DVESSADAT VVYNTIKFGA LDSTY
SEQ ID NO : 119	146197067	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	MLTUYFLLS LVVSLIEIGTO QSDEDHPKLTW QNGSSVSGS IVLDSNWRWV HDSGTTTNCYD GNJWSKDLCP SSNTSQKCY TEGDADYSGT GROSSGSKLT LRKFVTKGSYS TNIGSRVYL KDENTYEFPK LKNKEFPTV DDSKLNGGLN GALFYVAMDA DGGAKYKSSF KPGAKYGMGY CDAQCPHDMK FISGKANVDD WKPQDNDENS GNGKGTCGS EMDIWEGRM SQAYTVHACS KSGQYBCTGQ QCGDFTSGDR FKGTCDFDGK DYASWRGQDQ SFGYGRKTVD TKQPVTVYTO FICGDPLTEIR RLYVOGKTIK NMSKTSNLAD TYDSITDFKC DATKEASGDT NDFKAKGAMS GFSTNLNNGQ VLMWSLWDDH TANMLWLSDST YPTDSTKTGA SRGPCKAVSSG VPKDVEQSQHG DATVYSDIK FGAINSTFKW N
SEQ ID NO : 120	146197407	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	MLALALSIAK SLGIGTNAQE THDKLTKWRY QSKGSGCOTVN GEIVLDNSWV WTHHSCTNCY DGNTWTSLSLC PDPTCSNNC DLDDADYPTP YGISTSNSL KLGFVTHGSY STNIGSRVYL LRDSKNYMF KLNKEFPTF VDDSKLPCGL NGALIYVAMD EDDGVSNSI NKAGAQYGTG YCDAQCPHDN KFINGEANVL DMKPQSDNEN SGNGYGACC TEMPIWEANS MATAYTPHVC TVTGTRCEG TECGTDANQ RYNGICDGDG CDFNNSYPLGD KTFPGVKTV DSSEPVTVTT QFTNSNGQDS CTLSERTRKY VQGGKVIENS KUNLAGTAG NSVTDTFCNE OKKAGDND FEKGGLGKAL SKOLDAGNVL VLSLWDHHSV NMLWLDSTP TNAAAGALT ERGACATSSG APSDVEQSQP DATVYSDIK FGIDSTY
SEQ ID NO : 121	146197157	uncultured symbiotic protist of <i>Hodotermopsis sjostedti</i>	MLVIALILRG LSGVGTGTOQ EPHPSLWQ TSIGKSGCQSV SGSVYLDNSW RWTHTDTGTT NCYDGNEWS DLCPASTCS SNCGLEGADY SGTYGIGTSG SSILKLGDFVTK GSYSTNIGSR VYLLGDESHY KLFKLENNEF TFTVDDSNLE CGLNGALYEV AMEDDGASK YSGAKPCKAYI GMGYCDAQCP HDMKFINGDA NYEGWKSNSD DENATGKNG ACCEMDWE ANKATAYTPP HICTKNGEYR CEGTDGDTK DNRYGGYCD KDGCDFNISWR MGNQQFWGPG LIITIGKPTV VVTFPLADGK VQGGKVIENS KTGSMDEDF STIDFRCNQ KKAFFDTNDF EKRGGLKGIG TADGAVYLV LSLWDDHDVN MLWLDSTP DSGSKAGADR GPCATSSG KDVENSYASA SVTFSDIK FGIDSTY
SEQ ID NO : 122	146197403	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	MLLAFLFGK SLGIGTNAQE NHFRKLTWRY QSKGSGCOTVN GEIVLDNSWV WTHHSCTNCY DGNTWTSLSLC PDPTCSNNC DLDDADYPTP YGISTSNSL KLGFVTHGSY STNIGSRVYL LRDSKNYMF KLNKEFPTF VDDSKLPCGL NGALIYVAMD EDDGVSNSI NKAGAQYGTG YCDAQCPHDN KFINGEANVL DMKPQSDNEN SGNGYGACC TEMPIWEANS MATAYTPHVC TVTGTRCEG TECGTDANQ RYNGICDGDG CDFNNSYPLGD KSFPGVKTV DSSEPVTVTT QFTNSNGQDS CTLSERTRKY VQGGKVIENS KUNTAGMAG NSLTIDTRCN OKKAGDND FEKGGLGKAL SKOLDAGNVL VLSLWDHHSV NMLWLDSTP TNAAAGALT ERGACATSSG APSDVEQSQP DATVYSDIK FGIDSTY
SEQ ID NO : 123	146197081	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	MLASVYVLLS LVVSLIEIGTO QSEBHPKLTW QNGSSVSGS IVLDSNWRWV HDSGTTTNCYD GNJWSDDLCP NADTSSSKCY TEGDADYSGT GITSGSKVT LRKFVTKGSYS TNIGSRVYL KDENTYEFPK LKNKEFPTV DDSKLNGGLN GALFYVAMDA DGGAKYKSSF KPGAKYGMGY CDAQCPHDMK FISGKANVDD WKPQDNDENS DGKLGTCGS EMDIWEGRM SQAYTVHACS KSGQYBCTGQ QCGDFTSGDR FKGTCDFDGK DYASWRGQDQ

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 124	146197413	uncultured symbiotic protist of <i>Cryptococcus punctulatus</i>	SFYGRGKTVTD TKSPVTYVTO FICDPLTEIRI RYVVOGGKTI NSKTSNLAD TYDSITDKFCR DATKDATGDT NDFRKGKAMA GFSTNLNTQ VLSVHCMQ IOPICGLIR RIORIQKVV QAVDRVLCCR VFQRMVKASM VMLQSRTRL SLELSTRPLV GISPAGRLFF F
SEQ ID NO : 125	146197309	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	MILALLVLGK SLGIATNQAE THPLKTWTRY QSKGSGSTVN GEIVLDNSWR WTHSGTINCY DGTNTWTSLLC PDPPTCSNNC DLDAIDPQT YGLISTNSR YLGFTVHGSY STNIGSRVYL KLNKEEFTF VDDSSKLPCGLI NGAYEVANM ENGGVSNSI NKAGAQYGTG YCDACPHDM KFINGEANVL DKPKQSNDEN SGNGYGAAC TEMIWEANS MATAYTPHVC TVTGILERCEG TEGDIDNDQ RYNGICDKGD CDENSYRGD KSFFEVGKTV DSSEPVTVYT QFTTSNGODS GTLSEFRRKY VOGGKVIEVS KVNAVGTAG NSVTDTCNE QRKAGDNND FEKGGLGAL SKLDAGNL VLSLWDDHSV NMLWLDSTYP TNAAGALGT ERGACATSSG KPSDEQSQP DATYTESDIK FGIDSTY
SEQ ID NO : 126	146197227	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	MLCIGLISFY YSLQVGTNTA EPHPLKTMK GGQTYNGEVT VDSMRWHTH KGSITKNCYDG NLWSKDLCPD AATCGRNVCV EGAQYSGTIV RFTVTHGSYSTN VLSVHCMQ IOPICGLIR NCREFTTVD VSNLPGLNG ALYHVNMD EGTMKRYPDNE AGAKYGTGYC DAOCPDTLKK INGIPNSDGM KPOSNIDKNSG NGKYSGCCSE MDIWBNASIC SATVPHYCDN LQQTRCGTA CGENGGSRF GSSCDPDGCD FNSWRMGKNT FYGPBLIVDT KSKETVVTQF VGPVTEIKR KYVQNGKVE NSYNIEGMD KENVSVDERC TAQKKAFGDT DSFTTHGGFK QLGALAKGM VLSLWDDH TVNMLWDDV YPTNSKAGS DRGCPCPTSG VPADVESKSA DANVLYSDIR RGADSTYK
SEQ ID NO : 127	146197253	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	MLGALVALAS CIGGTGNTPE KHPDLKWTG GSIVSGSIVV DSNENRTHIK GETKNCYDGN LMSDKYCPDA ATCGRNVCV GADFSGTYGV TSQGSGDAATLK FVTHGQVSTN VGSRAIYLLK EKTYQMFNLV GREFTFTVDD SNLPGCLNG LYFQMDSDG CMKXPKDQNA CMKXPKDQYCD ACQDTDILK EKIPNDSWIK NGIPNSDGM KPOSNIDKNSG GKYGSCSEM DIWANSMAT AYTHVCDKL BCTRCGSSAC QONGGDRFS SS CDPDGCDF NSWRMGKNTF WGPGLIVDTK KPQVQVNTQFV GSGGSVTEIK RKYVQGGKV1 DNSMTNIAAM SKQYNSVSDF PCQAQKAFG DNDSTPKHGG FROGATLISK GHVVLVSLMD SVPTNSNKP GADRGPKTS SGVPSDVSEQ NADSTPKYSD IRFAGDSTY SK
SEQ ID NO : 128	146197099	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	MLAAALPFPA CSYVGVTKTT EPHPLKLMQZ CACKGCSQV SGEVMTDSW RWTHDNGKN CYDGNMVISI LCPDKTCSD KCVLDGAEFO ATGQIENGAT ALTPKRVTHG SYSTNSRSL YLUDKSTY VEQLNNEKFT FSVDSSKLPC GLMGALFYV MDGGGSKY AGAKPQAEVHG LGVCAQCPD SEGWKPSQGD KNAGCKYGS CCSMNDVYES NEONATALPH VCKTGTGOTRC SGKSRCGGOD GODRFAGNCF EDGCDPNUWR MGDKTFFGPP LTVDTKSPV VV7QFQGSPV TEIRRKVQN GKVIENAKS IPGIDATNAI SDTFCQOKK AFGDNTDFKN KGGETKLGSV FSGGMVLYLS LWDDHQYAML WLDSSTYPTNK DKSYPGVDRG PCPTSSCRPD DVESASGDAT VVYNTKFGA LDSTY
SEQ ID NO : 129	146197409	uncultured symbiotic protist of <i>Cryptococcus punctulatus</i>	MFGFLLSLFA LQFALIEIGTQ TSESHSPSTW EINGAROSQG IVIDSNWRML HDTSNWRML GNTWSSDLCP DPEKSONCY LEGADYSGY GSASGSGT LGFVTGGSY TNIGRVYLL KDENTPMFK LKNKEEFTF DVSNPFCGLN GALVFWAMEP DGAQEPHDW KFNGVTHGSY STNIGSRVYL KLNKEEFTF VDDSSKLPCGLI NGAYEVANM EDGEGTAKHSI I NKAGAQYGTG YCDACPHDM FNGEANVL WFGQSNIDEN GTGRGTCCT EMIDWEANQ ATTYVHACS KNARCGTEC GDDSAQSYRN GICDCKGDGF NSWRMGKNTF FGPGLIVDSS KPVTVTQPI GDPLTEIRI WVGGRKVN SFTNVSIGTS VDSTINTFCD ESKVATGDTN DEKAGGMSG FSKVLDTEVY LVLSLWDDHT ANMLWLDSTY PTDSTAIGAS RGCPATSSGD PRDVESASAN ASVKFSDIKF GALDSTY
SEQ ID NO : 130			MLASLPLSLN SLGIATNQAE THPLKTWTRY TGKAGQTVN GEIVLDNSWR WTHSGTINCY DGTNTWTSLLC PDPTCSNNC NLDOADYPT YGTTCSNOL KLGFTVHGSY STNIGSRVYL LRSKNYQMF KLNKEEFTF VDDSSKLPCGLI NGAYEVANM EDGEGTAKHSI I NKAGAQYGTG YCDACPHDM FNGEANVL WFGQSNIDEN SGNGYGAAC TEMIWEANS RAYATVHACS KNARCGTEC GTEGSDTNR YGGVCDKGDGK DFNSYRGD K SFFGQKTVD SSKEVTVVTO FITDNNQDSG KLTIEIRKVV QGGKVTDNSK

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 130	146197315	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	KKAEGDNDF EKRGIGLSA LG TOLEAGFVLY VLSLWDDHHSVN MLWLDSTYPT NASPGALGYE RGDCATSGV PADVQSQAD ASVTFSDIKF GPIDSTY
SEQ ID NO : 131	146197411	uncultured symbiotic protist of <i>Cryptococcus punctulatus</i>	MUCLIGLISFY YSLVGVTNTA ETHPKLTWNK GGQTNGEV TDSNMRWHT KGSTKNCYDG NLWSKDLCPD AATCCRNVCV EGAYISGTYG VTSNSGNALTL KFTVTHSYST NVGSPLYLK DEKTYQMPNL NGKEFTFTVD VSNLPQDLSG ALYVNGTGC DQCPIDK FQSQNDKNSG NGKYSGCCSE MDIWEMANSTC SAVTPHYCDN LQOTRQGAA CGENGGSRF GSSCDPDKF FNSWGMNKTF FYGPGLVDT KSKTIVVTOF VGPVTEIKR KVQNGKRVIE NSYNTIEGMD KENVSVDKFC TAQKKGATD DSFTKHGGFK OLGSALAKM VLWLSLWDDH TVNMLWLDSV YPTNSKKAGS DRGPCTPSG VPADVESKSA DANVYSDIR FGADSTYK
SEQ ID NO : 132	146197161	uncultured symbiotic Protist of <i>Hodotermopsis sjostedti</i>	MILALILVLGK SLGATNQAE THPLKTWRY QSKGSGSTVN GEIVLDSNWR WTHSGTNCY DGNTWSTSIC PDPPTCSNNC DLQDADYPT EGJTSSENSL KLGTVTHGSY STNTGSRVYL LDGSKNYMF KLNKEFTFTV VDDSLSPKGL NGAQYVAMN DPKQSDNIDEN SGNGRYGACC TEMDWEANS MATAYTHPHC TVTGELRRCG HDTKPFGNSA NYQDWKESDN KSFRFVGKTV DSSPRTVTVT QFTVNSNGDS GILSETRPKY VQGGKVIENS KYVAGITAG NSVTDTCNE QKKAGDND REKGIGLGL SKOLDAGNL VLSLWDDH SVNMLWLDSTP TNAAGALGT ERGACATSG KPSDYESQSP DATVTFSDIK FGIDSTY
SEQ ID NO : 133	146197323	uncultured symbiotic Protist of <i>Mastotermes darwiniensis</i>	MIGYLIQTYF FGIVGVTQOS ESHPSLWQQ CSKGGGTSV SGSTVLDSNW RWTHIPDGTT NCYDGNEWS DLCPPTCS NNCVLEGADY SGTYGISTSG SSAKLGFVTK GSYSTNIGSR VYLLGDESHY KLFDLRNKEP TFTVDSNLN CGLNGALYEV AMEDGGASR FTGLPKAGY GTGTYCDAQCP HDTKPFGNSA NYQDWKESDN DDNAGTGHY ACCTEMD IVE ANKATAYATF HICTENEYAR CEKSCGQDSS DDYXGGVCDK DCQDENSWRL GNQSWPGPL IIDIGKPVTVV VTOFVTDGTT DSGALEBIRR KYVOQGKTIIE NTUVKLSGID EVDSITDEFIC NQQRQAFDTI NDFEKKGHLS GLCKAFDYGV VLVLSLWDDH DVNMMLWLDSY YPTNPAGKAG ADRGPCATSS GDPKEVEDKY ASAATVTFSDI KRPIDSTY
SEQ ID NO : 134	146197077	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	MVFQGVFSFY YSIQVGVTNTA ETHPKLTWNK GGSTTNGEV TDSNMRWHT KGSTKNCYDG NLWSKDLCPD AATCCRNVCV EGAYISGTYG VTSNSGNALTL KFTVTHSYST NVGSPLYLK DEKTYQMPNL NGKEFTFTV VSQLPGGLN ALYVNGTGC DQCPIDK FQSQNDKNSG NGKYSGCCSE MDIWEMANSLA TAVTPHYCDQ VVOTRCEGRAS CGENGGDRF FWMGPGLVDT KSKTIVVTOF VGPVTEIKR EYVQGKRVIE NSYNTIEGMD KENVSVDKFC TAQKKGATD DSFTKHGGFS KLGSQFTKQ VLWLSLWDDH TVNMLWLDSV YPTNSKKAGS DRGPCTPSG DSSVYSDIR FGSTIDSTY
SEQ ID NO : 135	146197089	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	MLSFYFLGF GVSLLEIGTO SENHPTLSWQ QCTSSGSCSTS OSGSTVLDSEN WRWHDSGTT NCYDGNEWS DLCPPETCS KNCYLDGADY SGTYGITSNG SSSLKGEVTE GSYSTNIGSR VYLLGDESHY QIFKLKRHF TFTVYNSNLN CGLNGALYEV EMADGGKPK YPLARKGQY GMGTYCDAQCP HDMKFINGNA NYLDMKPOET DENSGERYG TCCFEMD IVE ANSQTATPF HICTENEYAR CEKSCGQDSS YEGEHTVDTK OPTIVTVOFI GDFLTETRRL YIQGGKVINN SKTONLASY FKARQAMAF SKNDLDPQVY VLWLWDDITA NMWLDSTP TDSRDATAER GPCATSSGPV KDVESNQDA SVVFEEDIKG AINSTSYN

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 136	146197091	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	MFGFLISLFA LQFALIEGTQ TSESHPSLTW EINGARSGQ IVIDSNWRWL HDSGTTTNCID GNTWSDILCP DPEKRSQNCY LEGADYSCTY GIASAGSGLT LGEVFTGSYS TWIGSRVYL KIDENTYOMFK LKNKEEFTTV DVSNLPGLN GALFVAMPS DGEAKYPLA KGAKYGMY CDAQCPHDWK FINGEANVLD WKFQSNNDNA GTGRGTCCT EMDIWEANQ ATTAYVHACS KNARCOTEC GDDSAQRYN GICDODGDF NSWRMGNKTIF FGPGLITVDSS KPVTIVTQPI GDLSTLWDTI RWTGVKIVN PFSNTVAGT VDSATNTFCD ESKVATGDPIN DEKAEGMSGM FSXALDTEVY LVLSLWDHT ANMLWIDSTY PNSNTAIGT RGPCATSSGD PKVNEASAN ASVKPSDIKF GARFSTY
SEQ ID NO : 137	146197097	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	MALVYFLLS LVVSLEIQTQ QSEDHPKLW QNGSSVSGS IVLDNSNRWV HDGSNTTNCYD GNJLWSTDLCI PSSDTCTSKCY TEGADYSCTY GITSSGSKVLT LKEVFTGSYS TWIGSRVYL KIDENTYEFPK LKNKEEFTTV DDSQNLCGLN GALFVAMDA DGEAKYTFSP KGAKYGMY CDAQCPHDWK FISCKANVDD WKFQSNNDNA GNGKLGTC CS EMDIWEENAK SQZTYVHACT KSGQYECTGQ QCGDTDSGDR FGKTCDCDGIC DYASMRWGQDQ SFYGBGKTVMD TKQPVTVVHQ FICDPLTEIR RLYVQGKTI INNSKTSNLAD TYDSITDIFC DATKEASGDT NDFKAKGAMS GFSMTNLNTAQ VLVLSLWDDH TANMLWLDST YPTDSTKTA SRGPACAVISG VPKDVESEQYG SAQVYVSDIK FGAINSTSY N
SEQ ID NO : 138	146197095	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	MALVYFLLS FVVELEIQTQ QSDDHPLKLTW QNGSSVSGS IVLDNSNRWV HDGSNTTNCYD GNJLWSTDLCI SSSDTSSKCY LEGADYSCTY GIASAGSGLT LKEVFTGSYS TWIGSRVYL KIDENTYEFPK LKGKEEFTTV DDSKLDCGLN GALFVAMDA DGEAKYTFSP KGAKYGMY CDAQCPHDWK FISCKANVDD WKFQSNNDNA GNGKLGTC CS EMDIWEENAK SQZTYVHACT KSGQYECTGQ QCGDTDSGDR FGKTCDCDGIC DYASMRWGQDQ SFYGBGKTVMD TKQPVTVVHQ FICDPLTEIR RLYVQGKTI INNSKTSNLAN VDTSITDIFC DUTKDASTG DATVYSDIK FGAINSTSY N
SEQ ID NO : 139	146197401	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	MFLALFVLGK SLGHATNOE NHPKLTWTRY QSKGSGQTVN GEVILDSNWR WTHHSNTGNYC DENTWSTSIC PDQTCSSNC DLDQADYDGT YGQSSGSNL KLGFTVHGSY STNIGSRVYL LRDUSKNTME KLNKEEFTTF VDDSSLPCLG NGALYFVAME EDGQVAKNSI NKAGAOYGTG YCDACCPHM KFINGEANVLD DKPQSNNDEN SGNGRYGACC TEMDWEANS MATAYTHVC TVTGIRCEG TECGDTDANQ RYNGICDKDG CDENSYRMGD KSFFEVGKTV DSSSPVTVUTV QFTTSNGDQ GTLSEKIRKY VQGGKVIENS KVNAGITAV NSITDTICNE QKKARGDNND FEKKGLGAL SKLDLGML VLSLWDDHSV MNLWLDSTP KPSDYESQSP DASTYFESDIK FGHDSTY
SEQ ID NO : 140	146197225	uncultured symbiotic protist of <i>Neoterpes koshunensis</i>	MLLCLSIAN SLGGTNTAE NHPKLSWNG GSVSQGTVY DANWENTHIK GETPNYCQDN LMSDKYCPDA ATCGKNCVIE GADYQFTGIV SSQGDLT FVTHQKSYT KTYQMNIN GKEFTFTVVD SNLPGLNG ALYHVNMDR GETKRYPDNE AGAKYGTGIC AQCPTDILKFI NGIPNSDVKP RQNDKNSGN GKYGSCSEM DIWEANSQAT AYTPHVCDSLK EQTRCSSSC GHGGERFS WGPGLIVDTK KPQVYVTOFV GSNSCPEIK RKYVQGKVI DNSMENIAGM SKOYNSVSDD FOQAOKKAFG DNDSPTRKGG FROQGATLKR GHVLVLSLWD DHDVNLWLD SVPTINSNKP GSDRGPCRTS SGIPADVESQ AASSSYKSD IRFGAIDSTK K
SEQ ID NO : 141	146197317	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	MLCIGLISFY YSLGGTNTA ETPHKLTWKN GGOTYNGEVT VDSNWRWHT KGSTKNCYDG NLWSKDLCPD AATCGKNCVIE EGAYISGTYG VTSQGNALIT KFTVTHSYST VNGSFLYLMK DEKTYOMENL NSKEFTFTVVD VSNUFLCGLN ALYHVNMDR GETKRYPDNE AGAKYGTGIC AQCPTDILKFI INGIPNSDVKP RQNDKNSGN NGKYSCCSE MDIBANSTIC SATPHVCDKL EQTRCSSSC GHGGERFS GSSCDPDGDF NSWRMGNKTIF FYGPGLIVDTK KSKEVVTVQF VGSPTVTELKR KYVQNGKVIEN NSFSNLEGMD DSFTKHGGPK QLGSGALARGM VLVLSLWDDH TNMLWLDSV YPTNSKAGS DRGPCTTSQ VPADVESKSA NANVYIYSDIR FGALDSTYK

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 142	146197251	uncultured symbiotic protist of <i>Neotermitessoshunensis</i>	MILCLGIGAS SLIDGNTIAE NHFOLSWNG GSSVSGSSTVU DANWNTIK GETKRNCDON IWSDRYCPDA ATCGCNCVIE GADFOGTGVY SAGSNALLT FVTGCGYSTN VGSRLYLLKD EKTYOIFULLI GKEFTFTVVDV SNLPGGLNGA LYTFQMDADG GTAKYSDNKA GAKYGTGYC ACQPTDLKFI INGIPNSDGMRK FOKNDKNSGN GRYGCCESEM DYNWANSLAT AYTPHVCDDKL ECQRCGRAC GONGEGDRFS SSCDPDGDF NSWRLENKTF WGPGLTVDTK QPVQVWTQV TKVQVQGKVI DNSFTKLDSL SCVPADEYESQ DNDSTPKHGG ERQGATLAK GHIVLISLWD DHDVNMWLQ SYVPTNSNKI GAIRGPCTIS SCVPADEYESQ ASSSVKYSK IRGAIDSTY K
SEQ ID NO : 143	146197319	uncultured symbiotic protist of <i>Mastotermitessdarwiniensis</i>	MLGIGEVCTV YSLVGIGNTTA ENHPKLTWN SGSTTINGEV TUDSNMRWHT KGJTKNYCDG NLWSKDLCPD AATCGCKNCVIL EGAYLSGTYG VNSPSGDALTL KFTVTHSYST NVGSPLYLJK DEXTYQIFNL NGKEFTFTVVD VSNLPGGLNG ALYVANMDAD GGTRGPDNO AGAKYGTGYC DAQCPTDLKFI INGIPNSDOW KPQSNDNSG NGKYGSCCSE MDTIFANSA TAWTPHYCDQ VGTQTRCEGRA CGENGGDRF GSSCDPDGCD FMSWRLENKTF FWGPGLIVDT KKPTVVTOTV YGGPVTETIKR KYVQGGRVIE NYSTYI EGLD KFNSISDGFIC TAQKKARGDN DSEFHGGFR OLGGSFTKQ VLWLSLMDDH TVMLWLDVS YPTNSKPGA DRGCPCTSSG VPADVESKNA GSSVYKSDIR FGSTDSTYK
SEQ ID NO : 144	146197071	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	MATLVGILVS LFALEVALDI GTOCTSEHPS LSWELINGQRO TGTSVTDNSW RWLHDGIFTN CYDGNEWSSD LCPDPBKCSQ NCYLEGADYS CTYGISSSGN SLOQEVGTVK SYSTEMGSRV TWLKDENTYFA TPKLKNKEFT FTADSNLPC GLNCALEYPA MPADGGSKY PLAKPGAKYQ MGYCAQCPH DMKTFINGEAN ILDWKP-SNDN ENAGAGRGT CCTIMDTEWA NSGATATVHV ACKSNARCEG TECGDDDRY NGICDKGCD FMSWRWGKD FFGPNLIVDS SKPTVVTOTP IGPDLTEIRR TYVQGGRVIO NSFTNISGVA SVDSITDAFC NENKVATGDT NDFKAKGCM SPSZALDTEV VLVLSLMDDH TANMLWLDST YPTSSALGA SPPCATTSQ EPKDVEZASA NASVFKPSDIK FGAIIDSTY
SEQ ID NO : 145	146197075	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	MLTLYVFLS LVSYEIGTO QSESHPOLSW QNGSSSVGS IVLDSNMRWV HDSGTTNCYD GNLSWSTDLCPS SSDTCTSKCY LEGDYSGY GITSHSGSLT LKFVTRGCVS YTSNRSVRL KDENTYEFK LKNKEFFTV DDSKLDGGLN GALFYVANDA DGKAKYTSFF KPGAKYGMGY CDAOCPHDMK FISCKANVDD WPOQDNDENS GNGKLGTCCS EMDIWEGNAK SQZYTVAHCT KSGQYECTGQ QCQGDTSGDR FKGTCDRDC DYAASWRWGQ SFYGGKTVD TKQBLTVYTO FVGDPLTEIR RVYVQGKTI INNSKTSNLAD TYDSITDGFIC DATKEAEGDT NDFKAKGAMS GPSMNLNTAQ VLWMSLMDDH TANMLWLDST YPTSTKTGA SRGPACAVSSG VPKDVEQSQHG DATVYSDIK FGAINSTPKW N
SEQ ID NO : 146	146197159	uncultured symbiotic protist of <i>Hodotermopsis sjostedti</i>	MLSLSPFLV GLGSESLGFT QOESHEHPS WONCSAKGSQ QSVSSIVLQ SMTWRLDHSQ TINCYDNEW STDPLCDAST CRDNCYIEGA DTSYDGTG SQAQLKQFV TKGYSTNIG SMTWRLDHSQ HQLFKLKH EFTFVYDDSQ LPCGSLNGALY FVMAEDGGA KPGAOYGMGY CDAOCPHDMK FITEGEANKD WPOQDNDENS GNGHGAACCT EMDIWENASQ ATYPYTHLCS KIGIYREGT ECGDANQR YNGCUDRG DENSYRUNK TFWGGLTVD SNKEMIVYTO FTTSNNODSG ELSETERIYV OGKTIONSD TNYQGITTN KITOAFCDET KVTFGDTINDF KAKGFGSGLS KSLESGAVLN LSWMDDHSVN MLWLDSTYPT DSAGKPGADR GPCAITSGDP KDVEEOSPNA SVTSDIKEG PIYSTY
SEQ ID NO : 147	146197405	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	MILALLVLGK SLGIAITNQAE THPKLTWTRY QSKGSGSTVN GEIVLDSNMR WTHHSGTNCY DGTNTWSTSIC PDPITCSNNC DLQDADYDGT YGISTSGERNL KLGFTVHGSY STNGTSRVL LKUTKSYENF KLKNKEFTFT VDDSTKUPCGI NGALYFEVAND EBGVKSNSI NKAGAQYGT YCDAOCPHDM KFINGEANVL DWKPOSUNEN SNGHGYACC TEMDWEANS MAYATPHVC TVTGURCEG TECGTDINDO RYNGICDKGD CDENSYRUGD KSFFEVGKTV DSSEPVTVYT QFTTSNGDS GTLSETERK VOGGKVIENS KVNVAGITAG NSVTDTCNE QRKAGDNND FERKGFGFAL SKOLVAGNL VLSLWDDHSV NMLWLDSTP KPSDVEQSQP DATVTEFDIK FGSDSTY

TABLE 1-continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO : 148	146197227	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	MICVQFLGLV YS1GIGINTQ EHPKLSWIKQ CSGGGCTTQ QGSVIVIDSAN RWIHSITDLT NCYDGNIWDS TLCPGTTCS KNCVLEGADY SGTYGTTSSG DSLTIKFVTH GSYSTNVGSR LYIILKDDNNY QIFNLACKEP TFTDVSNLP GLINGALYEV EMDQGGK K HNEAEGAKY GTGYCDAQCP DENSGENKGK SCCEMDIWE ANSLATAYTP HVCDTIGQKR COGTACGENG QGINKSEFWGPG LI1ITRKSVQ VVTFQFISGS SVEIREKVV QNGKVTEINY STLSGTEKIN SLMDHVVNM LMLDSVYPTN SNGPAGDGP CETSSGVPAD KKAEDTNSP ENHGEGKPRPS QHQDMVILVL SLMDHVVNM LMLDSVYPTN SNGPAGDGP CETSSGVPAD VESKASASV KYSDIRFGP1 DSTYK
SEQ ID NO : 149	146197261	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	MILCLWSTIAY SLGGTNTAB NHPKLSWKG GSSVSGSVTV DANWRWTHIK GETKNCYDGN LWSDKYCPDA ATCGKNCVIE GADFOGTGTVV SAGDGTLT FVTHGQYSTN VGSRLYLMKD EKTYQIENLN GHEFTFTVVD SNLPGCLNGA LYFQMDSPG GMKYGPDRQA GAKYGRGYCD ACQPITDLKFI NGIPNSDWK PKQNDKNSGN GKYGSCSEM DINFANSQAT AYTPHYCDKL ECTRCSGSAC GHTGGERFRS SSCDPDGDF NSWRMGNIKTF WGPGLIVDTK KPVOYVTQFV GSNSNCPBIK RRYVQCGKVI DNSMENIAGM TKQYNSVSDD FOQAQKKAFG DNDSTPKHGG FROUGATLGK GHVVLVLSWID DHDVNMWLWID SVYPTNSNP GSDRGPCKTS SGIPADVFSQ AASSZVKYSD IRGAIDSTY K

TABLE 2

Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Position corresponding to position 268	Position corresponding to position 411
SEQ ID NO: 1	BD29555*	Unknown	273	422
SEQ ID NO: 2	340514556	<i>Trichoderma reesei</i>	268	411
SEQ ID NO: 3	51243029	<i>Penicillium occitanis</i>	273	422
SEQ ID NO: 4	7cel (PDB) &	<i>Trichoderma reesei</i>	251	394
SEQ ID NO: 5	67516425	<i>Aspergillus nidulans</i> FGSC A4	274	424
SEQ ID NO: 6	46107376	<i>Gibberella zaeae</i> PH-1	268	415
SEQ ID NO: 7	70992391	<i>Aspergillus fumigatus</i> Af293	277	427
SEQ ID NO: 8	121699984	<i>Aspergillus clavatus</i> NRRL 1	277	427
SEQ ID NO: 9	1906845	<i>Claviceps purpurea</i>	269	416
SEQ ID NO: 10	1gpi (PDB) &	<i>Phanerochaete chrysosporium</i>	240	391
SEQ ID NO: 11	119468034	<i>Neosartorya fischeri</i> NRRL 181	265	414
SEQ ID NO: 12	7804883	<i>Leptosphaeria maculans</i>	256	401
SEQ ID NO: 13	85108032	<i>Neurospora crassa</i> N150	268	412
SEQ ID NO: 14	169859458	<i>Coprinopsis cinerea</i> okayama	270	421
SEQ ID NO: 15	154292161	<i>Botryotinia fuckeliana</i> B05-10	—	410
SEQ ID NO: 16	169615761 #	<i>Phaeosphaeria nodorum</i> SN15	246	393
SEQ ID NO: 17	4883502	<i>Humicola grisea</i>	272	413
SEQ ID NO: 18	950686	<i>Humicola grisea</i>	270	416
SEQ ID NO: 19	124491660	<i>Chaetomium thermophilum</i>	272	413
SEQ ID NO: 20	58045187	<i>Chaetomium thermophilum</i>	270	416
SEQ ID NO: 21	169601100 #	<i>Phaeosphaeria nodorum</i> SN15	237	383
SEQ ID NO: 22	169870197	<i>Coprinopsis cinerea</i> okayama	269	421
SEQ ID NO: 23	3913806	<i>Agaricus bisporus</i>	263	414
SEQ ID NO: 24	169611094	<i>Phaeosphaeria nodorum</i> SN15	270	414
SEQ ID NO: 25	3131	<i>Phanerochaete chrysosporium</i>	—	410
SEQ ID NO: 26	70991503	<i>Aspergillus fumigatus</i> Af293	265	414
SEQ ID NO: 27	294196	<i>Phanerochaete chrysosporium</i>	258	409
SEQ ID NO: 28	18997123	<i>Thermoascus aurantiacus</i>	268	418
SEQ ID NO: 29	4204214	<i>Humicola grisea</i> var <i>thermoidea</i>	272	413
SEQ ID NO: 30	34582632	<i>Trichoderma viride</i> (also known as <i>Hypocheira rufa</i> )	268	411
SEQ ID NO: 31	156712284	<i>Thermoascus aurantiacus</i>	268	418
SEQ ID NO: 32	39977899	<i>Magnaporthe grisea</i> ( <i>oryzae</i> ) 70-15	268	414
SEQ ID NO: 33	20986705	<i>Talaromyces emersonii</i>	266	416
SEQ ID NO: 34	22138843	<i>Aspergillus oryzae</i>	265	414
SEQ ID NO: 35	55775695	<i>Penicillium chrysogenum</i>	276	426
SEQ ID NO: 36	171676762	<i>Podospora anserina</i>	270	417
SEQ ID NO: 37	146350520	<i>Pleurotus</i> sp Florida	268	420
SEQ ID NO: 38	37732123	<i>Gibberella zeae</i>	268	415
SEQ ID NO: 39	156055188	<i>Sclerotinia sclerotiorum</i> 1980	—	410
SEQ ID NO: 40	453224	<i>Phanerochaete chrysosporium</i>	258	409
SEQ ID NO: 41	50402144	<i>Trichoderma reesei</i>	268	411
SEQ ID NO: 42	115397177	<i>Aspergillus terreus</i> NIH2624	274	424
SEQ ID NO: 43	154312003	<i>Botryotinia fuckeliana</i> B05-10	266	416
SEQ ID NO: 44	49333365	<i>Volvariella volvacea</i>	268	420
SEQ ID NO: 45	729650	<i>Penicillium janthinellum</i>	274	424
SEQ ID NO: 46	146424871	<i>Pleurotus</i> sp Florida	267	418
SEQ ID NO: 47	67538012	<i>Aspergillus nidulans</i> FGSC A4	265	410
SEQ ID NO: 48	62006162	<i>Fusarium poae</i>	268	415
SEQ ID NO: 49	146424873	<i>Pleurotus</i> sp Florida	267	418
SEQ ID NO: 50	295937	<i>Trichoderma viride</i>	268	411
SEQ ID NO: 51	6179889 #	<i>Alternaria alternata</i>	240	386
SEQ ID NO: 52	119483864	<i>Neosartorya fischeri</i> NRRL 181	278	428
SEQ ID NO: 53	85083281	<i>Neurospora crassa</i> OR74A	270	412
SEQ ID NO: 54	3913803	<i>Cryphonectria parasitica</i>	269	416
SEQ ID NO: 55	60729633	<i>Corticium rolfsii</i>	265	415
SEQ ID NO: 56	39971383	<i>Magnaporthe grisea</i> 70-15	268	410
SEQ ID NO: 57	39973029	<i>Magnaporthe grisea</i> 70-15	269	410
SEQ ID NO: 58	1170141	<i>Fusarium oxysporum</i>	268	415
SEQ ID NO: 59	121710012	<i>Aspergillus clavatus</i> NRRL 1	265	414
SEQ ID NO: 60	17902580	<i>Penicillium funiculosum</i>	273	422
SEQ ID NO: 61	1346226	<i>Humicola grisea</i> var <i>thermoidea</i>	270	416
SEQ ID NO: 62	156712282	<i>Chaetomium thermophilum</i>	270	416
SEQ ID NO: 63	169768818	<i>Aspergillus oryzae</i> RIB40	277	427
SEQ ID NO: 64	46241270	<i>Gibberella pulicaris</i>	268	415
SEQ ID NO: 65	49333363	<i>Volvariella volvacea</i>	265	418
SEQ ID NO: 66	46395332	<i>Irpex lacteus</i>	263	414
SEQ ID NO: 67	50844407 #	<i>Chaetomium thermophilum</i> var <i>thermophilum</i>	245	391
SEQ ID NO: 68	4586347	<i>Irpex lacteus</i>	264	415
SEQ ID NO: 69	3980202	<i>Phanerochaete chrysosporium</i>	258	410
SEQ ID NO: 70	27125837	<i>Melanocarpus albomyces</i>	273	414
SEQ ID NO: 71	171696102	<i>Podospora anserina</i>	265	415
SEQ ID NO: 72	3913802	<i>Cochliobolus carbonum</i>	270	416
SEQ ID NO: 73	50403723	<i>Trichoderma viride</i>	268	411

TABLE 2-continued

Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Position corresponding to position 268	Position corresponding to position 411
SEQ ID NO: 74	3913798	<i>Aspergillus aculeatus</i>	275	425
SEQ ID NO: 75	66828465	<i>Dictyostelium discoideum</i>	269	419
SEQ ID NO: 76	156060391	<i>Sclerotinia sclerotiorum</i> 1980	252	402
SEQ ID NO: 77	116181754	<i>Chaetomium globosum</i> CBS 148-51	263	413
SEQ ID NO: 78	145230535	<i>Aspergillus niger</i>	274	424
SEQ ID NO: 79	46241266	<i>Nectria haematococca</i> mpVI	268	415
SEQ ID NO: 80	1q9h (PDB) #	<i>Talaromyces emersonii</i>	248	398
SEQ ID NO: 81	157362170	<i>Polyporus arcularius</i>	269	420
SEQ ID NO: 82	7804885	<i>Leptosphaeria maculans</i>	267	407
SEQ ID NO: 83	121852	<i>Phanerochaete chrysosporium</i>	258	409
SEQ ID NO: 84	126013214	<i>Penicillium decumbens</i>	264	415
SEQ ID NO: 85	156048578	<i>Sclerotinia sclerotiorum</i> 1980	265	413
SEQ ID NO: 86	156712278	<i>Acremonium thermophilum</i>	269	414
SEQ ID NO: 87	21449327	<i>Aspergillus nidulans</i>	265	410
SEQ ID NO: 88	171683762	<i>Podospora anserina</i>	274	415
SEQ ID NO: 89	56718412	<i>Thermoascus aurantiacus</i> var <i>levisporus</i>	268	418
SEQ ID NO: 90	15824273	<i>Pseudotrichonympha grassii</i>	263	414
SEQ ID NO: 91	115390801	<i>Aspergillus terreus</i> NIH2624	266	411
SEQ ID NO: 92	453223	<i>Phanerochaete chrysosporium</i>	258	409
SEQ ID NO: 93	3132	<i>Phanerochaete chrysosporium</i>	—	407
SEQ ID NO: 94	16304152	<i>Thermoascus aurantiacus</i>	268	417
SEQ ID NO: 95	156712280	<i>Acremonium thermophilum</i>	273	420
SEQ ID NO: 96	5231154	<i>Volvariella volvacea</i>	281	438
SEQ ID NO: 97	116200349	<i>Chaetomium globosum</i> CBS 148-51	270	412
SEQ ID NO: 98	4586343	<i>Irpex lacteus</i>	263	414
SEQ ID NO: 99	15321718	<i>Lentinula edodes</i>	—	417
SEQ ID NO: 100	146424875	<i>Pleurotus</i> sp <i>Florida</i>	267	418
SEQ ID NO: 101	62006158	<i>Fusarium venenatum</i>	268	415
SEQ ID NO: 102	296027	<i>Phanerochaete chrysosporium</i>	258	409
SEQ ID NO: 103	154449709	<i>Fusicoccum</i> sp BCC4124	272	424
SEQ ID NO: 104	169859460	<i>Coprinopsis cinerea</i> okayama	269	421
SEQ ID NO: 105	50400675	<i>Trichoderma harzianum</i>	264	407
SEQ ID NO: 106	729649	<i>Neurospora crassa</i>	262	406
SEQ ID NO: 107	119472134	<i>Neosartorya fischeri</i> NRRL 181	277	427
SEQ ID NO: 108	117935080	<i>Chaetomium thermophilum</i>	272	413
SEQ ID NO: 109	154300584	<i>Botryotinia fuckeliana</i> B05-10	265	413
SEQ ID NO: 110	15824271	<i>Pseudotrichonympha grassii</i>	263	414
SEQ ID NO: 111	4586345	<i>Irpex lacteus</i>	263	414
SEQ ID NO: 112	46241268	<i>Gibberella avenacea</i>	268	416
SEQ ID NO: 113	6164684	<i>Aspergillus niger</i>	274	424
SEQ ID NO: 114	6164682	<i>Aspergillus niger</i>	266	412
SEQ ID NO: 115	33733371	<i>Chrysosporium lucknowense</i> US6573086-10	269	415
SEQ ID NO: 116	29160311	<i>Thielavia australiensis</i>	269	415
SEQ ID NO: 117	146197087	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	260	402
SEQ ID NO: 118	146197237	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	264	409
SEQ ID NO: 119	146197067	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	260	402
SEQ ID NO: 120	146197407	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	261	412
SEQ ID NO: 121	146197157	uncultured symbiotic protist of <i>Hodotermopsis sjostedti</i>	264	410
SEQ ID NO: 122	146197403	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	261	412
SEQ ID NO: 123	146197081	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	260	410
SEQ ID NO: 124	146197413	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	261	412
SEQ ID NO: 125	146197309	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	259	402
SEQ ID NO: 126	146197227	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	258	404
SEQ ID NO: 127	146197253	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	264	409
SEQ ID NO: 128	146197099	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	258	401
SEQ ID NO: 129	146197409	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	260	411
SEQ ID NO: 130	146197315	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	259	402

TABLE 2-continued

Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Position corresponding to position 268	Position corresponding to position 411
SEQ ID NO: 131	146197411	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	261	412
SEQ ID NO: 132	146197161	uncultured symbiotic protist of <i>Hodotermopsis sjostedti</i>	263	413
SEQ ID NO: 133	146197323	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	259	402
SEQ ID NO: 134	146197077	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	264	415
SEQ ID NO: 135	146197089	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	258	400
SEQ ID NO: 136	146197091	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	258	401
SEQ ID NO: 137	146197097	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	260	402
SEQ ID NO: 138	146197095	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	260	402
SEQ ID NO: 139	146197401	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	261	412
SEQ ID NO: 140	146197225	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	258	404
SEQ ID NO: 141	146197317	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	259	402
SEQ ID NO: 142	146197251	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	258	404
SEQ ID NO: 143	146197319	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	259	402
SEQ ID NO: 144	146197071	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	259	402
SEQ ID NO: 145	146197075	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	260	402
SEQ ID NO: 146	146197159	uncultured symbiotic protist of <i>Hodotermopsis sjostedti</i>	260	410
SEQ ID NO: 147	146197405	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	261	412
SEQ ID NO: 148	146197327	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	264	408
SEQ ID NO: 149	146197261	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	258	404

TABLE 3

SEQ ID NO:	Database Accession Number	Species of Origin	Signal sequence (SS) start and end position	Catalytic Domain (CD) start and end position	Linker start and end position	Cellulose Binding Domain (CBD) start and end
SEQ ID NO: 1	BD29555*	Unknown	1-25	26-455	456-493	494-529
SEQ ID NO: 2	340514556	<i>Trichoderma reesei</i>	1-17	18-444	445-479	480-514
SEQ ID NO: 3	51243029	<i>Penicillium occitanis</i>	1-25	26-455	456-493	494-529
SEQ ID NO: 4	7cel (PDB) &	<i>Trichoderma reesei</i>	N/A	1-427	N/A	N/A
SEQ ID NO: 5	67516425	<i>Aspergillus nidulans</i> FGSC A4	1-23	24-457	458-490	491-526
SEQ ID NO: 6	46107376	<i>Gibberella zeae</i> PH-1	1-17	18-448	449-476	477-512
SEQ ID NO: 7	70992391	<i>Aspergillus fumigatus</i> Af293	1-26	27-460	461-496	497-532
SEQ ID NO: 8	121699984	<i>Aspergillus clavatus</i> NRRL 1	1-27	27-460	461-503	504-539
SEQ ID NO: 9	1906845	<i>Claviceps purpurea</i>	1-19	20-449	N/A	N/A
SEQ ID NO: 10	1gpi (PDB) &	<i>Phanerochaete chrysosporium</i>	N/A	1-424	N/A	N/A
SEQ ID NO: 11	119468034	<i>Neosartorya fischeri</i> NRRL 181	1-17	18-447	N/A	N/A
SEQ ID NO: 12	7804883	<i>Leptosphaeria maculans</i>	1-17	18-434	N/A	N/A
SEQ ID NO: 13	85108032	<i>Neurospora crassa</i> N150	1-17	18-445	446-485	486-521
SEQ ID NO: 14	169859458	<i>Coprinopsis cinerea</i> <i>okayama</i>	1-18	19-454	N/A	N/A
SEQ ID NO: 15	154292161	<i>Botryotinia fuckeliana</i> B05-10	1-18	19-443	444-555	556-596
SEQ ID NO: 16	169615761 #	<i>Phaeosphaeria nodorum</i> SN15	1	2-426	N/A	N/A

TABLE 3-continued

SEQ ID NO:	Database Accession Number	Species of Origin	Signal sequence (SS) start and end position	Catalytic Domain (CD) start and end position	Linker start and end position	Cellulose Binding Domain (CBD) start and end
SEQ ID NO: 17	4883502	<i>Humicola grisea</i>	1-22	23-446	N/A	N/A
SEQ ID NO: 18	950686	<i>Humicola grisea</i>	1-18	19-449	450-489	490-525
SEQ ID NO: 19	124491660	<i>Chaetomium thermophilum</i>	1-22	23-446	N/A	N/A
SEQ ID NO: 20	58045187	<i>Chaetomium thermophilum</i>	1-18	19-449	450-494	495-530
SEQ ID NO: 21	169601100 #	<i>Phaeosphaeria nodorum</i> SN15	1	2-416	N/A	N/A
SEQ ID NO: 22	169870197	<i>Coprinopsis cinerea</i> okayama	1-18	19-454	N/A	N/A
SEQ ID NO: 23	3913806	<i>Agaricus bisporus</i>	1-18	19-447	448-470	471-506
SEQ ID NO: 24	169611094	<i>Phaeosphaeria nodorum</i> SN15	1-18	19-447	N/A	N/A
SEQ ID NO: 25	3131	<i>Phanerochaete chrysosporium</i>	1-19	20-443	N/A	N/A
SEQ ID NO: 26	70991503	<i>Aspergillus fumigatus</i> Af293	1-17	18-447	N/A	N/A
SEQ ID NO: 27	294196	<i>Phanerochaete chrysosporium</i>	1-18	19-442	443-480	481-516
SEQ ID NO: 28	18997123	<i>Thermoascus aurantiacus</i>	1-17	18-451	N/A	N/A
SEQ ID NO: 29	4204214	<i>Humicola grisea</i> var <i>thermoidea</i>	1-22	23-446	N/A	N/A
SEQ ID NO: 30	34582632	<i>Trichoderma viride</i> (also known as <i>Hypocheira rufa</i> )	1-18	18-444	445-479	480-514
SEQ ID NO: 31	156712284	<i>Thermoascus aurantiacus</i>	1-17	18-451	N/A	N/A
SEQ ID NO: 32	39977899	<i>Magnaporthe grisea</i> ( <i>oryzae</i> ) 70-15	1-17	18-447	N/A	N/A
SEQ ID NO: 33	20986705	<i>Talaromyces emersonii</i>	1-18	19-449	N/A	N/A
SEQ ID NO: 34	22138843	<i>Aspergillus oryzae</i>	1-17	18-447	N/A	N/A
SEQ ID NO: 35	55775695	<i>Penicillium chrysogenum</i>	1-25	26-459	460-494	495-529
SEQ ID NO: 36	171676762	<i>Podospora anserina</i>	1-18	19-450	451-492	493-528
SEQ ID NO: 37	146350520	<i>Pleurotus</i> sp <i>Florida</i>	1-18	19-453	N/A	N/A
SEQ ID NO: 38	37732123	<i>Gibberella zae</i>	1-17	18-448	449-476	477-512
SEQ ID NO: 39	156055188	<i>Sclerotinia sclerotiorum</i> 1980	1-18	19-443	444-546	547-586
SEQ ID NO: 40	453224	<i>Phanerochaete chrysosporium</i>	1-18	19-442	443-474	475-510
SEQ ID NO: 41	50402144	<i>Trichoderma reesei</i>	1-17	18-444	445-478	479-513
SEQ ID NO: 42	115397177	<i>Aspergillus terreus</i> NIH2624	1-23	24-457	458-505	506-541
SEQ ID NO: 43	154312003	<i>Botryotinia fuckeliana</i> B05-10	1-17	18-449	450-480	481-516
SEQ ID NO: 44	49333365	<i>Volvariella volvacea</i>	1-18	19-453	N/A	N/A
SEQ ID NO: 45	729650	<i>Penicillium janthinellum</i>	1-25	26-456	457-502	503-537
SEQ ID NO: 46	146424871	<i>Pleurotus</i> sp <i>Florida</i>	1-18	19-451	452-487	488-523
SEQ ID NO: 47	67538012	<i>Aspergillus nidulans</i> FGSC A4	1-17	18-443	N/A	N/A
SEQ ID NO: 48	62006162	<i>Fusarium poae</i>	1-17	18-448	449-475	476-511
SEQ ID NO: 49	146424873	<i>Pleurotus</i> sp <i>Florida</i>	1-18	19-451	452-487	488-523
SEQ ID NO: 50	295937	<i>Trichoderma viride</i>	1-17	18-444	445-478	479-513
SEQ ID NO: 51	6179889 #	<i>Alternaria alternata</i>	1	2-419	N/A	N/A
SEQ ID NO: 52	119483864	<i>Neosartorya fischeri</i> NRRL 181	1-26	27-461	462-499	500-535
SEQ ID NO: 53	85083281	<i>Neurospora crassa</i> OR74A	1-20	21-445	N/A	N/A
SEQ ID NO: 54	3913803	<i>Cryphonectria parasitica</i>	1-18	19-449	N/A	N/A
SEQ ID NO: 55	60729633	<i>Corticium rolfsii</i>	1-18	19-448	449-492	493-528
SEQ ID NO: 56	39971383	<i>Magnaporthe grisea</i> 70-15	1-17	18-443	N/A	N/A
SEQ ID NO: 57	39973029	<i>Magnaporthe grisea</i> 70-15	1-19	20-443	N/A	N/A
SEQ ID NO: 58	1170141	<i>Fusarium oxysporum</i>	1-17	18-448	449-478	479-514
SEQ ID NO: 59	121710012	<i>Aspergillus clavatus</i> NRRL 1	1-17	18-447	N/A	N/A
SEQ ID NO: 60	17902580	<i>Penicillium funiculosum</i>	1-25	26-455	456-493	494-529

TABLE 3-continued

SEQ ID NO:	Database Accession Number	Species of Origin	Signal sequence (SS) start and end position	Catalytic Domain (CD) start and end position	Linker start and end position	Cellulose Binding Domain (CBD) start and end
SEQ ID NO: 61	1346226	<i>Humicola grisea</i> var <i>thermoidea</i>	1-18	19-449	450-489	490-525
SEQ ID NO: 62	156712282	<i>Chaetomium thermophilum</i>	1-18	19-449	450-496	497-532
SEQ ID NO: 63	169768818	<i>Aspergillus oryzae</i> RIB40	1-25	26-460	N/A	N/A
SEQ ID NO: 64	46241270	<i>Gibberella pulicaris</i>	1-17	18-448	449-474	475-510
SEQ ID NO: 65	49333363	<i>Volvariella volvacea</i>	1-18	19-451	452-476	477-512
SEQ ID NO: 66	46395332	<i>Irpex lacteus</i>	1-18	19-447	448-485	486-521
SEQ ID NO: 67	50844407 #	<i>Chaetomium thermophilum</i> var <i>thermophilum</i>	N/A	1-424	425-469	470-505
SEQ ID NO: 68	4586347	<i>Irpex lacteus</i>	1-18	19-448	449-490	491-526
SEQ ID NO: 69	3980202	<i>Phanerochaete chrysosporium</i>	1-18	19-443	444-475	476-511
SEQ ID NO: 70	27125837	<i>Melanocarpus albonymces</i>	1-23	23-447	N/A	N/A
SEQ ID NO: 71	171696102	<i>Podospora anserina</i>	1-17	17-448	N/A	N/A
SEQ ID NO: 72	3913802	<i>Cochliobolus carbonum</i>	1-18	19-449	N/A	N/A
SEQ ID NO: 73	50403723	<i>Trichoderma viride</i>	1-17	18-444	445-479	480-514
SEQ ID NO: 74	3913798	<i>Aspergillus aculeatus</i>	1-22	23-458	459-505	506-540
SEQ ID NO: 75	66828465	<i>Dictyostelium discoideum</i>	1-19	20-452	N/A	N/A
SEQ ID NO: 76	156060391	<i>Sclerotinia sclerotiorum</i> 1980	1-17	18-435	436-470	471-504
SEQ ID NO: 77	116181754	<i>Chaetomium globosum</i> CBS 148-51	1-17	18-446	N/A	N/A
SEQ ID NO: 78	145230535	<i>Aspergillus niger</i>	1-21	22-457	458-500	501-536
SEQ ID NO: 79	46241266	<i>Nectria haematococca</i> mpVI	1-18	18-448	449-472	473-508
SEQ ID NO: 80	1q9h (PDB) #	<i>Talaromyces emersonii</i>	N/A	1-431	N/A	N/A
SEQ ID NO: 81	157362170	<i>Polyporus arcularius</i>	1-18	19-453	N/A	N/A
SEQ ID NO: 82	7804885	<i>Leptosphaeria maculans</i>	1-20	21-440	N/A	N/A
SEQ ID NO: 83	121852	<i>Phanerochaete chrysosporium</i>	1-18	19-442	443-480	481-516
SEQ ID NO: 84	126013214	<i>Penicillium decumbens</i>	1-17	18-448	N/A	N/A
SEQ ID NO: 85	156048578	<i>Sclerotinia sclerotiorum</i> 1980	1-16	17-446	N/A	N/A
SEQ ID NO: 86	156712278	<i>Acremonium thermophilum</i>	1-17	18-447	448-487	488-523
SEQ ID NO: 87	21449327	<i>Aspergillus nidulans</i>	1-17	18-443	N/A	N/A
SEQ ID NO: 88	171683762	<i>Podospora anserina</i>	1-22	23-448	N/A	N/A
SEQ ID NO: 89	56718412	<i>Thermoascus aurantiacus</i> var <i>levisporus</i>	1-17	18-451	N/A	N/A
SEQ ID NO: 90	15824273	<i>Pseudotrichonympha grassii</i>	1-20	21-447	N/A	N/A
SEQ ID NO: 91	115390801	<i>Aspergillus terreus</i> NIH2624	1-17	18-444	N/A	N/A
SEQ ID NO: 92	453223	<i>Phanerochaete chrysosporium</i>	1-18	19-442	443-474	475-510
SEQ ID NO: 93	3132	<i>Phanerochaete chrysosporium</i>	1-19	20-436	437-467	468-504
SEQ ID NO: 94	16304152	<i>Thermoascus aurantiacus</i>	1-17	18-450	N/A	N/A
SEQ ID NO: 95	156712280	<i>Acremonium thermophilum</i>	1-21	22-453	N/A	N/A
SEQ ID NO: 96	5231154	<i>Volvariella volvacea</i>	1-15	16-472	473-500	501-536
SEQ ID NO: 97	116200349	<i>Chaetomium globosum</i> CBS 148-51	1-20	21-445	N/A	N/A
SEQ ID NO: 98	4586343	<i>Irpex lacteus</i>	1-18	19-447	448-481	482-517
SEQ ID NO: 99	15321718	<i>Lentinula edodes</i>	1-18	19-450	451-480	481-516
SEQ ID NO: 100	146424875	<i>Pleurotus sp Florida</i>	1-18	19-451	452-487	488-523
SEQ ID NO: 101	62006158	<i>Fusarium venenatum</i>	1-17	18-448	449-471	472-507
SEQ ID NO: 102	296027	<i>Phanerochaete chrysosporium</i>	1-18	19-442	443-480	481-516

TABLE 3-continued

SEQ ID NO:	Database Accession Number	Species of Origin	Signal sequence (SS) start and end position	Catalytic Domain (CD) start and end position	Linker start and end position	Cellulose Binding Domain (CBD) start and end
SEQ ID NO: 103	154449709	<i>Fusicoccum</i> sp BCC4124	1-19	20-457	N/A	N/A
SEQ ID NO: 104	169859460	<i>Coprinopsis cinerea</i> okayama	1-18	19-454	N/A	N/A
SEQ ID NO: 105	50400675	<i>Trichoderma harzianum</i>	1-17	18-440	441-470	471-505
SEQ ID NO: 106	729649	<i>Neurospora crassa</i>	1-17	18-439	440-480	481-516
SEQ ID NO: 107	119472134	<i>Neosartorya fischeri</i> NRRL 181	1-26	27-460	461-494	495-530
SEQ ID NO: 108	117935080	<i>Chaetomium thermophilum</i>	1-22	23-446	N/A	N/A
SEQ ID NO: 109	154300584	<i>Botryotinia fuckeliana</i> B05-10	1-16	17-446	N/A	N/A
SEQ ID NO: 110	15824271	<i>Pseudorichonympha grassii</i>	1-20	21-447	N/A	N/A
SEQ ID NO: 111	4586345	<i>Ipxex lacteus</i>	1-18	19-447	448-487	488-523
SEQ ID NO: 112	46241268	<i>Gibberella avenacea</i>	1-17	18-449	450-478	478-513
SEQ ID NO: 113	6164684	<i>Aspergillus niger</i>	1-21	22-457	458-500	501-536
SEQ ID NO: 114	6164682	<i>Aspergillus niger</i>	1-17	18-445	N/A	N/A
SEQ ID NO: 115	33733371	<i>Chrysosporium lucknowense</i> US6573086-10	1-17	18-448	449-490	491-526
SEQ ID NO: 116	29160311	<i>Thielavia australiensis</i>	1-18	18-448	449-502	503-538
SEQ ID NO: 117	146197087	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-22	23-435	N/A	N/A
SEQ ID NO: 118	146197237	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	1-20	21-442	N/A	N/A
SEQ ID NO: 119	146197067	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-22	23-435	N/A	N/A
SEQ ID NO: 120	146197407	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	1-19	20-445	N/A	N/A
SEQ ID NO: 121	146197157	uncultured symbiotic protist of <i>Hodotermopsis sjostedti</i>	1-20	21-443	N/A	N/A
SEQ ID NO: 122	146197403	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	1-19	20-445	N/A	N/A
SEQ ID NO: 123	146197081	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-22	23-443	N/A	N/A
SEQ ID NO: 124	146197413	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	1-19	20-445	N/A	N/A
SEQ ID NO: 125	146197309	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	1-20	21-435	N/A	N/A
SEQ ID NO: 126	146197227	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	1-19	20-437	N/A	N/A
SEQ ID NO: 127	146197253	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	1-21	21-442	N/A	N/A
SEQ ID NO: 128	146197099	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-22	23-434	N/A	N/A
SEQ ID NO: 129	146197409	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	1-19	20-444	N/A	N/A

TABLE 3-continued

SEQ ID NO:	Database Accession Number	Species of Origin	Signal sequence (SS) start and end position	Catalytic Domain (CD) start and end position	Linker start and end position	Cellulose Binding Domain (CBD) start and end
SEQ ID NO: 130	146197315	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	1-20	21-435	N/A	N/A
SEQ ID NO: 131	146197411	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	1-19	20-445	N/A	N/A
SEQ ID NO: 132	146197161	uncultured symbiotic protist of <i>Hodotermopsis sjostedti</i>	1-20	21-446	N/A	N/A
SEQ ID NO: 133	146197323	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	1-20	21-435	N/A	N/A
SEQ ID NO: 134	146197077	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-21	22-448	N/A	N/A
SEQ ID NO: 135	146197089	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-22	23-433	N/A	N/A
SEQ ID NO: 136	146197091	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-22	23-434	N/A	N/A
SEQ ID NO: 137	146197097	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-22	23-435	N/A	N/A
SEQ ID NO: 138	146197095	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-22	23-435	N/A	N/A
SEQ ID NO: 139	146197401	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	1-19	20-445	N/A	N/A
SEQ ID NO: 140	146197225	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	1-19	20-437	N/A	N/A
SEQ ID NO: 141	146197317	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	1-20	21-435	N/A	N/A
SEQ ID NO: 142	146197251	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	1-19	20-437	N/A	N/A
SEQ ID NO: 143	146197319	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	1-20	21-435	N/A	N/A
SEQ ID NO: 144	146197071	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-25	26-435	N/A	N/A
SEQ ID NO: 145	146197075	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-22	23-435	N/A	N/A
SEQ ID NO: 146	146197159	uncultured symbiotic protist of <i>Hodotermopsis sjostedti</i>	1-23	24-443	N/A	N/A
SEQ ID NO: 147	146197405	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	1-19	20-445	N/A	N/A
SEQ ID NO: 148	146197327	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	1-20	21-441	N/A	N/A

TABLE 3-continued

SEQ ID NO:	Database Accession Number	Species of Origin	Signal sequence (SS) start and end position	Catalytic Domain (CD) start and end position	Linker start and end position	Cellulose Binding Domain (CBD) start and end
SEQ ID NO: 149	146197261	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	1-19	20-437	N/A	N/A

TABLE 4

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in active site loop residues in sequence identifier	Amino acid positions of fragment in active site loop residues in sequence identifier	Position of catalytic residue in sequence identifier
SEQ ID NO : 150 BD29555*		Unknown	NVEGWT <del>PSS</del> NNANT <u>G</u> GNH <del>G</del> ACC <u>A</u> E <u>D</u> IV <u>M</u> ANS	210-242	214-226	234, 239
SEQ ID NO : 151 340514556	Trichoderma reesei		NVEGWT <del>PSS</del> NNANT <u>G</u> GNH <del>G</del> ACC <u>A</u> E <u>D</u> IV <u>M</u> ANS	205-237	209-221	229, 234
SEQ ID NO : 152 51243029	Penicillium occitanis		NVEGWT <del>PSS</del> NNANT <u>G</u> GHGS <u>C</u> <del>S</del> MDIV <u>M</u> ANS	210-242	214-226	234, 239
SEQ ID NO : 153 7cel (PDB) & Trichoderma reesei			NVEGWT <del>PSS</del> NNANT <u>G</u> GHGS <u>C</u> <del>S</del> MDIV <u>M</u> ANS	188-220	192-204	212, 217
SEQ ID NO : 154 67516425	Aspergillus nidulans FGSC A4		NVEGWT <u>S</u> DT <u>P</u> NG <u>G</u> GNH <u>S</u> CCA <u>E</u> NDIV <u>M</u> ANS	211-243	215-227	235, 240
SEQ ID NO : 155 46107376	Gibberella zaeae PH-1		NSDGW <u>O</u> P <u>D</u> SD <u>Y</u> NG <u>G</u> GNL <u>G</u> TCC <u>P</u> EMD <u>M</u> ANS	205-237	209-221	229, 234
SEQ ID NO : 156 70992391	Aspergillus fumigatus Af293		NVEGWT <u>O</u> P <u>S</u> SD <u>Y</u> NA <u>G</u> T <u>G</u> HN <u>H</u> SCCA <u>E</u> MDIV <u>M</u> ANS	214-246	218-230	238, 243
SEQ ID NO : 157 121699984	Aspergillus clavatus NRRL 1		NVEGWT <u>P</u> SS <u>S</u> SD <u>Y</u> NA <u>G</u> N <u>G</u> HN <u>H</u> SCCA <u>E</u> MDIV <u>M</u> ANS	214-246	218-230	238, 243
SEQ ID NO : 158 1906845	Claviceps purpurea		NSKDWT <u>P</u> SK <u>S</u> SD <u>Y</u> NA <u>G</u> I <u>G</u> S <u>L</u> G <u>A</u> CC <u>R</u> <del>E</del> MDIV <u>M</u> ANN	206-238	210-222	230, 235
SEQ ID NO : 159 19pi (PDB) & Phanerochaete chrysosporium			NVGNWT <u>E</u> T <u>G</u> --SNT <u>G</u> T <u>G</u> S <u>G</u> TCC <u>C</u> <del>E</del> MDIV <u>M</u> ANN	185-215	189-199	207, 212
SEQ ID NO : 160 119468034	Neosartorya fischeri NRRL 181		NVEGWT <del>P</del> SS <u>S</u> DN <u>K</u> NA <u>G</u> V <u>G</u> HN <u>H</u> SCC <u>C</u> <del>E</del> MDIV <u>M</u> ANS	202-234	206-218	226, 231
SEQ ID NO : 161 7804883	Leptosphaeria maculans		NVEGWT <u>O</u> P <u>S</u> RD <u>Q</u> NA <u>G</u> V <u>G</u> HN <u>H</u> SCCA <u>E</u> MDIV <u>M</u> ANS	193-225	197-209	217, 222
SEQ ID NO : 162 85108032	Neurospora crassa N150 (OR74A)		NVEGWT <u>P</u> ST <u>D</u> NA <u>A</u> G <u>D</u> H <u>G</u> T <u>C</u> <del>C</del> MDIV <u>M</u> ANK	205-237	209-221	229, 234
SEQ ID NO : 163 169859458	Coprinopsis cinerea okayama		NSADWT <u>P</u> SE <u>D</u> PN <u>A</u> G <u>R</u> GR <u>Y</u> GI <u>C</u> CA <u>E</u> MDIV <u>M</u> ANS	207-239	211-223	231, 236
SEQ ID NO : 164 154292161	Botryotinia fuckeliana B05-10		NVEGWT <u>P</u> D <u>S</u> NS <u>A</u> S <u>G</u> T <u>G</u> NI <u>G</u> SCC <u>C</u> <del>E</del> FDIV <u>M</u> ANS	203-235	207-219	227, 232

TABLE 4 -continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of active site loop residues in sequence identifier	Amino acid positions of active site loop residues in sequence identifier	Position of catalytic residues in sequence identifier
SEQ ID NO : 165 169615761 #		Phaeosphaeria nodorum SN15	NADG <b>WQASTSDPNA</b> G <b>V</b> KKGACCA <b>E</b> MDV <b>R</b> ANS	183-215	187-199	207, 212
SEQ ID NO : 166 4883502		Humicola grisea	NIE <b>GWPSTNDPNA</b> G <b>V</b> YGP <b>M</b> ACCA <b>E</b> LDV <b>R</b> ESNA	208-240	212-224	232, 237
SEQ ID NO : 167 950686		Humicola grisea	NIE <b>GWTGSTNDPNA</b> G <b>V</b> RYGT <b>C</b> CE <b>E</b> MDV <b>R</b> ANN	207-239	211-223	231, 236
SEQ ID NO : 168 124491660		Chaetomium thermophilum	NIE <b>GWPSTNDPNA</b> G <b>V</b> YGP <b>M</b> ACCA <b>E</b> LDV <b>R</b> ESNA	209-241	213-225	233, 238
SEQ ID NO : 169 58045187		Chaetomium thermophilum	NIE <b>NWPSTNDPNA</b> G <b>F</b> GR <b>Y</b> GS <b>C</b> SE <b>E</b> MDV <b>R</b> ANN	207-239	211-223	231, 236
SEQ ID NO : 170 169601100 #		Phaeosphaeria nodorum SN15	NVE <b>GWPSPDNDPNA</b> G <b>V</b> GHGS <b>C</b> CA <b>E</b> MDV <b>R</b> ANS	174-206	178-190	198, 203
SEQ ID NO : 171 169870197		Coprinopsis cinerea okyama	NSVG <b>WEPSETDSNAGR</b> GR <b>Y</b> GI <b>C</b> CA <b>E</b> MDV <b>R</b> ANS	207-239	211-223	231, 236
SEQ ID NO : 172 3913806		Agaricus bisporus	NSEG <b>WGPSPDNAG</b> AT <b>G</b> FG <b>A</b> CG <b>M</b> DI <b>D</b> V <b>R</b> ANS	203-235	207-219	227, 232
SEQ ID NO : 173 169611094		Phaeosphaeria nodorum SN15	NVE <b>GWPSPDADPNAGSK</b> KG <b>A</b> CC <b>P</b> MD <b>D</b> V <b>R</b> ANS	208-240	212-224	232, 237
SEQ ID NO : 174 3131		Phanerochaete chrysosporium	NVQ <b>GWNAT</b> --AT <b>TG</b> TG <b>S</b> Y <b>G</b> SC <b>C</b> CT <b>E</b> LD <b>D</b> V <b>R</b> ANS	204-234	208-218	226, 231
SEQ ID NO : 175 70991503		Aspergillus fumigatus Af293	NVE <b>GWPSSDKNA</b> G <b>V</b> GHGS <b>C</b> CP <b>E</b> MDV <b>R</b> ANS	202-234	206-218	226, 231
SEQ ID NO : 176 294196		Phanerochaete chrysosporium	NVE <b>GWNTA</b> --AN <b>A</b> T <b>G</b> T <b>Y</b> GT <b>C</b> CT <b>E</b> MDV <b>R</b> ANN	203-233	207-217	225, 230
SEQ ID NO : 177 18997123		Thermoascus aurantiacus	NVE <b>GWP</b> SANDPNA <b>G</b> V <b>G</b> HGS <b>C</b> CA <b>E</b> MDV <b>R</b> ANS	205-237	209-221	229, 234
SEQ ID NO : 178 4204214		Humicola grisea var thermocidea	NIE <b>GWPSTNDPNA</b> G <b>V</b> YGP <b>M</b> ACCA <b>E</b> LDV <b>R</b> ESNA	208-240	212-224	232, 237
SEQ ID NO : 179 34582632		Trichoderma viride (also known as Hypochrea rufa)	NVE <b>GWPSSNNANTG</b> GHGS <b>C</b> SE <b>E</b> MDV <b>R</b> ANS	205-237	209-221	229, 234
SEQ ID NO : 180 156712284		Thermoascus aurantiacus	NVE <b>GWP</b> SANDPNA <b>G</b> V <b>G</b> HGS <b>C</b> CA <b>E</b> MDV <b>R</b> ANS	205-237	209-221	229, 234

## US 9,096,871 B2

97

98

TABLE 4 -continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in active site loop residues in sequence identifier	Amino acid positions of active site loop residues in sequence identifier	Position of catalytic residues in sequence identifier
SEQ ID NO : 181 39977899	Magnaporthe grisea (oryzae)	NVEGW <del>PSS</del> DANS <del>G</del> YGNMGS <del>CC</del> A <del>E</del> MDIV <del>R</del> ANS 70-15	205-237	209-221	229,	234
SEQ ID NO : 182 20986705	Talaromyces emersonii	NVEGW <del>PSS</del> NNNNTG <del>G</del> DHGSCCA <del>E</del> MDIV <del>R</del> ANS	203-235	207-219	227,	232
SEQ ID NO : 183 22138843	Aspergillus oryzae	R-KGWPSPSD <del>D</del> XNA <del>G</del> YGHGSCCP <del>O</del> MDIV <del>R</del> ANS	203-234	206-218	226,	231
SEQ ID NO : 184 55775695	Penicillium chrysogenum	NVEGW <del>PSS</del> DYNGT <del>G</del> NYGS <del>CC</del> A <del>E</del> MDIV <del>R</del> ANS	213-245	217-229	237,	242
SEQ ID NO : 185 171676762	Podospora anserina	NIEGNP <del>S</del> TNDVNAGA <del>G</del> RYGTCC <del>E</del> MDIV <del>R</del> ANN	207-239	211-223	231,	236
SEQ ID NO : 186 146350520	Pleurotus sp Florida	NVQGW <del>PSS</del> NDNSA <del>G</del> KGQXGS <del>CC</del> A <del>E</del> MDIV <del>R</del> ANS	207-239	211-223	231,	236
SEQ ID NO : 187 37732123	Gibberella zeae	NSDGW <del>PSS</del> PDIVNGTG <del>G</del> LNGLTC <del>C</del> PFMDIV <del>R</del> ANS	205-237	209-221	229,	234
SEQ ID NO : 188 15605188	Sclerotinia sclerotiorum 1980	NNEGWP <del>D</del> SNSANS <del>G</del> TGNIGS <del>CC</del> SEFDIV <del>R</del> ANS	203-235	207-219	227,	232
SEQ ID NO : 189 453224	Phanerochaete chrysosporium	NVGNWTETG--SNTG <del>G</del> SYGTCC <del>E</del> MDIV <del>R</del> ANN	203-233	207-217	225,	230
SEQ ID NO : 190 50402144	Trichoderma reesei	NVEGW <del>PSS</del> NNNNTG <del>G</del> GHGS <del>CC</del> E <del>M</del> DIV <del>R</del> ANS	205-237	209-221	229,	234
SEQ ID NO : 191 115397177	Aspergillus terreus NIH2624	NVEGP <del>S</del> ANDANAG <del>G</del> QXGS <del>CC</del> A <del>E</del> MDIV <del>R</del> ANS	211-243	215-227	235,	240
SEQ ID NO : 192 154312003	Botryotinia fuckeliana B05-10	NSVGW <del>T</del> P <del>S</del> NDVNAGA <del>G</del> QXGS <del>CC</del> A <del>E</del> MDIV <del>R</del> ANK	206-238	210-222	230,	235
SEQ ID NO : 193 49333365	Volvariella volvacea	NVQGW <del>PSS</del> NDVNAGT <del>G</del> NYGACCNEMDV <del>R</del> ANS	207-239	211-223	231,	236
SEQ ID NO : 194 729650	Penicillium janthinellum	NVDG <del>W</del> T <del>P</del> SKNDVN <del>G</del> HNHS <del>CC</del> A <del>E</del> MDIV <del>R</del> ANS	211-243	215-227	235,	240
SEQ ID NO : 195 146424871	Pleurotus sp Florida	NILDWSASATDANA <del>G</del> RYGACCA <del>E</del> MDIV <del>R</del> ANS	206-238	210-222	230,	235

TABLE 4 -continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of active site loop residues in sequence identifier	Amino acid positions of active site loop residues in sequence identifier	Position of catalytic residue in sequence identifier
SEQ ID NO : 196 67538012		Aspergillus nidulans FGSC A4	NVEGWEPSSDSDANAGYGVGMGTCCP <del>E</del> MDIV <del>R</del> ANS	202-234	206-218	226, 231
SEQ ID NO : 197 62006162		Fusarium poae	NSDGWEP <del>S</del> KSDYNG <del>S</del> GNLGTC <del>C</del> P <del>E</del> MDIV <del>R</del> ANS	205-237	209-221	229, 234
SEQ ID NO : 198 146424873		Pleurotus sp Florida	NILDNGS <del>A</del> TDPNAG <del>N</del> GRYACCA <del>E</del> MDIV <del>R</del> ANS	206-238	210-222	230, 235
SEQ ID NO : 199 295937		Trichoderma viride	NVEGWEP <del>S</del> NNANTGIGHGSCC <del>E</del> MDIV <del>R</del> ANS	205-237	209-221	229, 234
SEQ ID NO : 200 6179889 #		Alternaria alternata	NVEGWKPSSNDANAGYGVGHGS <del>C</del> CA <del>E</del> MDIV <del>R</del> ANS	177-209	181-193	201, 206
SEQ ID NO : 201 119483864		Neosartorya fischeri NRRL 181	NVEGWTP <del>S</del> NNNTGIGNYGS <del>C</del> CA <del>E</del> MDIV <del>R</del> NSN	215-247	219-231	239, 244
SEQ ID NO : 202 85083281		Neurospora crassa OR74A	NIEGWTP <del>S</del> DANAGYGVGYGCCA <del>E</del> IDV <del>R</del> NSNA	207-239	211-223	231, 236
SEQ ID NO : 203 3913803		Cryphonectria parasitica	NVEGWTP <del>S</del> IDANAGYVG <del>L</del> GSCC <del>E</del> MDIV <del>R</del> ANS	206-238	210-222	230, 235
SEQ ID NO : 204 60729633		Corticium rolfsii	NLLDNWATS- <del>A</del> NS <del>G</del> TG <del>S</del> YSCC <del>P</del> MDIV <del>R</del> ANK	206-236	210-220	228, 233
SEQ ID NO : 205 39971383		Magnaporthe grisea 70-15	NIEGWOP <del>S</del> TDSSAGIGAQGCCA <del>E</del> IDIV <del>R</del> SNK	205-237	209-221	229, 234
SEQ ID NO : 206 39973029		Magnaporthe grisea 70-15	NIEGWKPSSNDANAGYGVGYGCCA <del>E</del> IDV <del>R</del> NSNA	206-238	210-222	230, 235
SEQ ID NO : 207 1170141		Fusarium oxysporum	NSEGWKPSSDSDYNAGYGVNLGTC <del>C</del> P <del>E</del> MDIV <del>R</del> ANS	205-237	209-221	229, 234
SEQ ID NO : 208 121710012		Aspergillus clavatus NRRL 1	NVEGWEP <del>S</del> ND <del>D</del> KNAGYGVGYGS <del>C</del> CP <del>E</del> MDIV <del>R</del> ANS	202-234	206-218	226, 231
SEQ ID NO : 209 17902580		Penicillium funiculosum	NVEGWTP <del>S</del> NNNTGIGNYGS <del>C</del> CA <del>E</del> MDIV <del>R</del> ANS	210-242	214-226	234, 239
SEQ ID NO : 210 1346226		Humicola grisea var thermotropaea	NIEGTWTG <del>S</del> NDPNAGAGRYGTC <del>C</del> SEMDIV <del>R</del> ANN	207-239	211-223	231, 236

TABLE 4 -continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in active site loop sequence identifier	Amino acid positions of active site loop residues in sequence identifier	Position of catalytic residues in sequence identifier
SEQ ID NO : 211 156712282		Chaetomium thermophilum	NVGNWTPSTNDANAGFGRYGS <del>CCSEMDIV</del> <u>RANN</u>	207-239	211-223	231, 236
SEQ ID NO : 212 169768818	RIB40	Aspergillus oryzae	NVEGWWSSTNNNANTGTGNHGCCA <u>EIDIV</u> ESNS	214-246	218-230	238, 243
SEQ ID NO : 213 46241270	Gibberella pulicaris		NSDGWQPSKSDYNAGTGNMGTCCP <u>EIDIV</u> RANS	205-237	209-221	229, 234
SEQ ID NO : 214 49333363	Volvariella volvacea		NVAGVNGSPPDNTNA <u>GTC</u> GNMAGCC <u>EIDIV</u> RANS	205-237	209-221	229, 234
SEQ ID NO : 215 46395332	Irpea lacteus		NVAGHTGSSSDPNSGTYGNYGTC <u>EIDIV</u> RANS	202-234	206-218	226, 231
SEQ ID NO : 216 50844407 #	Chaetomium thermophilum var thermophilum		NIENWTPSTNDANAGFGRYGS <del>CCSEMDIV</del> <u>RANN</u>	182-214	186-198	206, 211
SEQ ID NO : 217 4586347	Irpea lacteus		NIVDWTASGDANS <u>GTC</u> SPGTC <u>EIDIV</u> RANS	203-235	207-219	227, 232
SEQ ID NO : 218 3980202	Phanerochaete chrysosporium		NVGENWTPG--SNTGTGSGTGTCC <u>EIDIV</u> <u>RANN</u>	203-233	207-217	225, 230
SEQ ID NO : 219 27125837	Melanocarpus albomyces		NIEGWWSTS <u>D</u> PNAGVGYGPyGS <u>EIDIV</u> ESNA	210-242	214-226	234, 239
SEQ ID NO : 220 171696102	Podospora anserina		NVEGGGGAD--GNSGTGKGYGI <u>EIDIV</u> RANS	206-236	210-220	228, 233
SEQ ID NO : 221 3913802	Cochliobolus carbonum		NVEGNWPS <u>D</u> ADPNG <u>GAK</u> KIACCP <u>EIDIV</u> RANS	208-240	212-224	232, 237
SEQ ID NO : 222 50403723	Trichoderma viride		NVEGWEP <u>S</u> STNNNANTGTGHGCC <u>EIDIV</u> RANS	205-237	209-221	229, 234
SEQ ID NO : 223 3913798	Aspergillus aculeatus		NIEGWEPS <u>S</u> TDYNAGTGNHGCCP <u>EIDIV</u> RANS	210-242	214-226	234, 239
SEQ ID NO : 224 66828465	Dictyostelium discoideum		NVDGWP <u>S</u> TNNPNTG <u>G</u> LNLSGCC <u>EIDIV</u> <u>RANN</u>	206-238	210-222	230, 235
SEQ ID NO : 225 156060391	Sclerotinia sclerotiorum 1980		NSVGWTP <u>S</u> SDNDYNTG <u>Q</u> YGS <u>CCSEMDIV</u> <u>RANK</u>	192-224	196-208	216, 221

TABLE 4 -continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of active site loop residues in sequence identifier	Amino acid positions of active site loop residues in sequence identifier	Position of catalytic residues in sequence identifier
SEQ ID NO : 226 116181754		Chaetomium globosum CBS 148-51	NSE <span style="text-decoration: underline;">GEGGED</span> --GN <span style="text-decoration: underline;">SGTGKYGTC</span> CCAEMDIVRANL	203-233	207-217	225, 230
SEQ ID NO : 227 145230535		Aspergillus niger	N <span style="text-decoration: underline;">DGWEPSSNNNNTG</span> DHGSCCAEMDV <span style="text-decoration: underline;">RANS</span>	209-241	213-225	233, 238
SEQ ID NO : 228 46241266		Nectria haematococca mpVI	NSDEWKPS <span style="text-decoration: underline;">SDSDKNAGY</span> GKYGTC <span style="text-decoration: underline;">CP</span> EMDIVRANK	205-237	209-221	229, 234
SEQ ID NO : 229 1q9n (PDB) #		Talaromyces emersonii	NVE <span style="text-decoration: underline;">GWPSSNNNANTG</span> DHGSCCAEMDV <span style="text-decoration: underline;">RANS</span>	185-217	189-201	209, 214
SEQ ID NO : 230 157362170		Polyporus arcularius	NVLD <span style="text-decoration: underline;">WAGSSNDPNA</span> GTHGHGTCCCNMDIV <span style="text-decoration: underline;">RANS</span>	208-240	212-224	232, 237
SEQ ID NO : 231 7804885		Leptosphaeria maculans	NAE <span style="text-decoration: underline;">GWTKSASD</span> DPNSGYGKKGACCAQMDIV <span style="text-decoration: underline;">RANS</span>	204-236	208-220	228, 233
SEQ ID NO : 232 121852		Phanerochaete chrysosporium	NVE <span style="text-decoration: underline;">GWNATSA</span> -ANA <span style="text-decoration: underline;">GHTGNYCCT</span> EMDIVRANN	203-233	207-217	225, 230
SEQ ID NO : 233 126013214		Penicillium decumbens	NVE <span style="text-decoration: underline;">GWPKSAND</span> KNAGYGP <span style="text-decoration: underline;">HGSCCAEMDIV</span> RANS	201-233	205-217	225, 230
SEQ ID NO : 234 156048578		Sclerotinia sclerotiorum 1980	NVD <span style="text-decoration: underline;">GWWPSSNNNPTG</span> NYGSCCAEMDIV <span style="text-decoration: underline;">RANS</span>	202-234	206-218	226, 231
SEQ ID NO : 235 156712278		Acromonium thermophilum	NIDGWP <span style="text-decoration: underline;">SSNDANAGL</span> GNHGSCCSEMDIVRANK	206-238	210-222	230, 235
SEQ ID NO : 236 21449327		Aspergillus nidulans (also known as Emericella nidulans)	NVE <span style="text-decoration: underline;">GWPSSD</span> DANAGY <span style="text-decoration: underline;">G</span> MGTCCP <span style="text-decoration: underline;">EMDIV</span> RANS	202-234	206-218	226, 231
SEQ ID NO : 237 171683762		Podospora anserina (S mat+)	NIE <span style="text-decoration: underline;">GRESSNDENAGY</span> GPGGCCAEMDIV <span style="text-decoration: underline;">RNSNA</span>	211-243	215-227	235, 240
SEQ ID NO : 238 56718412		Thermoascus aurantiacus var leviporus	NVE <span style="text-decoration: underline;">GWPSSANDPNA</span> GYNHGSCCAE <span style="text-decoration: underline;">MDIV</span> RANS	205-237	209-221	229, 234
SEQ ID NO : 239 158244273		Pseudotrichonympha grassii	NVENWKPO <span style="text-decoration: underline;">TNDENAGNGRY</span> ACCTEMDIVRANK	200-232	204-216	224, 229

## US 9,096,871 B2

105

106

TABLE 4 -continued

Sequence Identifier (SEQ ID No.)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in active site loop residues in sequence identifier	Amino acid positions of active site loop residues in sequence identifier	Position of catalytic residues in sequence identifier
SEQ ID NO : 240 115390801		<i>Aspergillus terreus</i> NIH2624	NVEGWTTPSDIDND <b>KNA</b> G <b>VGHGS</b> CCP <b>E</b> MDIV <b>WANS</b>	203-235	207-219	227, 232
SEQ ID NO : 241 452223		<i>Phanerochaete chrysosporium</i>	NVGNW <b>TETG</b> --SNT <b>GTSY</b> GTCC <b>EMD</b> IV <b>WANN</b>	203-233	207-217	225, 230
SEQ ID NO : 242 3132		<i>Phanerochaete chrysosporium</i>	NVEGWL <b>GTT</b> --ATTG <b>TGFF</b> SCCT <b>DIA</b> LM <b>FAND</b>	202-232	206-216	224, 229
SEQ ID NO : 243 16304152		<i>Thermoascus aurantiacus</i>	NVEG <b>WPSA</b> ND <b>PNA</b> G <b>YGN</b> G <b>HGS</b> CA <b>E</b> MDIV <b>WANS</b>	205-237	209-221	229, 234
SEQ ID NO : 244 156712280		<i>Acremonium thermophilum</i>	NSASWOPSSND <b>QNA</b> G <b>YGY</b> GM <b>GSCCA</b> <b>E</b> MDIV <b>WANS</b>	210-242	214-226	234, 239
SEQ ID NO : 245 5231154		<i>Volvariella volvacea</i>	NVQ <b>GWPSP</b> ND <b>TNA</b> G <b>TG</b> NYG <b>ACCN</b> R <b>MDIV</b> <b>WANS</b>	220-252	224-236	244, 249
SEQ ID NO : 246 116200349		<i>Chaetomium globosum</i> CBS 148-51	NYDGWTTPSSND <b>ANAG</b> YG <b>ALG</b> GG <b>CC</b> <b>E</b> MDIV <b>WNSNA</b>	207-239	211-223	231, 236
SEQ ID NO : 247 4586343		<i>Irpex lacteus</i>	NVAG <b>WAGS</b> ASD <b>DPNAGS</b> GT <b>LGT</b> CC <b>EMD</b> IV <b>WANN</b>	202-234	206-218	226, 231
SEQ ID NO : 248 15321718		<i>Lentinula edodes</i>	NVEGWTTPSS <b>SPNAGT</b> GT <b>GIC</b> C <b>EMD</b> IV <b>WANS</b>	208-240	212-224	232, 237
SEQ ID NO : 249 146424875		<i>Pleurotus sp</i> Florida	NVLDWSASATDD <b>NAG</b> RYG <b>ACCA</b> <b>E</b> MDIV <b>WANS</b>	206-238	210-222	230, 235
SEQ ID NO : 250 62006158		<i>Fusarium venenatum</i>	NSDG <b>WOPSKD</b> YNG <b>ST</b> GNL <b>GTC</b> CP <b>E</b> MDIV <b>WANS</b>	205-237	209-221	229, 234
SEQ ID NO : 251 296027		<i>Phanerochaete chrysosporium</i>	NVEG <b>WMA</b> T--ANAG <b>TG</b> NYG <b>TC</b> C <b>EMD</b> IV <b>WANN</b>	203-233	207-217	225, 230
SEQ ID NO : 252 154449709		<i>Fusicoccum sp</i> BCC4124	NVQNNTASS <b>TD</b> <b>KNA</b> G <b>TG</b> HYG <b>GS</b> CC <b>N</b> <b>EMD</b> IV <b>WANS</b>	209-241	213-225	233, 238
SEQ ID NO : 253 169859460		<i>Coprinopsis cinerea</i> okayama	NSVGWEPE <b>SET</b> DPNAG <b>KQ</b> GY <b>GIC</b> CA <b>E</b> MDIV <b>WANS</b>	207-239	211-223	231, 236
SEQ ID NO : 254 50400675		<i>Trichoderma harzianum</i> (anamorph of <i>Hypocrea lixii</i> )	NVEGWEPESSNN <b>ANT</b> GYG <b>GHGS</b> CC <b>E</b> MDIV <b>WANS</b>	201-233	205-217	225, 230

TABLE 4 -continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of active site loop residues in sequence identifier	Amino acid positions of active site loop residues in sequence identifier	Position of catalytic residues in sequence identifier
SEQ ID NO : 255 729649	Neurospora crassa (OR74A)		NVEGWT <u>P</u> STNDAN-GIGDHGS <u>C</u> SE <u>M</u> DIV <u>E</u> ANK	200-231	204-215	223, 228
SEQ ID NO : 256 119472134	<i>Neosartoria Fischeri</i> NRRL 181		NVEGWP <u>S</u> SSNDANAGT <u>G</u> YGS <u>C</u> CA <u>E</u> MDIV <u>E</u> ANS	214-246	218-230	238, 243
SEQ ID NO : 257 11793505080	<i>Chaetomium thermophilum</i>		NIEGWP <u>R</u> STNDANAGYGPY <u>G</u> ACCA <u>E</u> MDIV <u>E</u> NSNA	209-241	213-225	233, 238
SEQ ID NO : 258 154300584	<i>Botryotinia fuckeliana</i> B05-10		NVDGWP <u>W</u> P <u>S</u> SSNNT <u>G</u> YGNHGS <u>C</u> CA <u>E</u> MDIV <u>E</u> ANS	202-234	206-218	226, 231
SEQ ID NO : 259 15824271	<i>Pseudotrichonympha ha Grassii</i>		NVENIKWP <u>O</u> TNDENAG <u>N</u> GRY <u>G</u> ACCT <u>E</u> MDIV <u>E</u> ANK	200-232	204-216	224, 229
SEQ ID NO : 260 4586345	<i>Irpea lacteus</i>		NVEGWT <u>G</u> SS <u>T</u> DS <u>N</u> SG <u>T</u> GN <u>G</u> TC <u>C</u> SE <u>M</u> DIV <u>E</u> ANS	202-234	206-218	226, 231
SEQ ID NO : 261 46241268	<i>Gibberella avenacea</i>		NSDGWP <u>P</u> SD <u>D</u> INAG <u>I</u> GN <u>M</u> GTCCP <u>E</u> NDIV <u>E</u> ANS	205-237	209-221	229, 234
SEQ ID NO : 262 6164684	<i>Aspergillus niger</i>		NDCGWEP <u>S</u> SSNN <u>V</u> NT <u>G</u> YGDHGSC <u>C</u> AE <u>M</u> DIV <u>E</u> ANS	209-241	213-225	233, 238
SEQ ID NO : 263 6164682	<i>Aspergillus niger</i>		NVDGWP <u>P</u> SS <u>N</u> ND <u>N</u> T <u>G</u> HN <u>H</u> GS <u>C</u> CP <u>E</u> MDIV <u>E</u> ANK	203-235	207-219	227, 232
SEQ ID NO : 264 33733371	<i>Chrysosporium lucknowense</i> US6573086-10		NVENWP <u>S</u> STNDANAGT <u>G</u> K <u>G</u> SC <u>C</u> SE <u>M</u> DIV <u>E</u> ANN	206-238	210-222	230, 235
SEQ ID NO : 265 29160311	<i>Thielavia australiensis</i>		NVEGWP <u>S</u> STNDANAGS <u>G</u> K <u>G</u> SC <u>C</u> CT <u>E</u> MDIV <u>E</u> ANN	206-238	210-222	230, 235
SEQ ID NO : 266 146197087	uncultured symbiotic protist of <i>Reticulitermes speratus</i>		NVDDW <u>W</u> K <u>P</u> Q <u>D</u> ND <u>E</u> NS <u>G</u> KL <u>G</u> TC <u>C</u> SE <u>M</u> DIV <u>E</u> GNM	197-229	201-213	221, 226
SEQ ID NO : 267 146197237	uncultured symbiotic protist of <i>Neotermes koshuensis</i>		NSEGWP <u>S</u> SD <u>K</u> NA <u>G</u> NG <u>K</u> YGS <u>C</u> SE <u>M</u> DIV <u>E</u> NS	200-232	204-216	224, 229
SEQ ID NO : 268 146197067	uncultured symbiotic protist of <i>Reticulitermes speratus</i>		NVDDW <u>W</u> K <u>P</u> Q <u>D</u> ND <u>E</u> NS <u>G</u> KL <u>G</u> TC <u>C</u> SE <u>M</u> DIV <u>E</u> GNM	197-229	201-213	221, 226

TABLE 4 -continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in active site loop residues in sequence identifier	Amino acid positions of active site loop residues in sequence identifier	Position of catalytic residue in sequence identifier
SEQ ID NO : 269	146197407	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	NVLDWPKPQSDNDENSGNGRYGACCTEMDIV <sub>■</sub> ANS	198-230	202-214	222, 227
SEQ ID NO : 270	146197157	uncultured symbiotic protist of <i>Hodotermopsis sjostedti</i>	NVEGWKPQSDNDENAGTGYKGACCTEMDIV <sub>■</sub> RANK	201-233	205-217	225, 230
SEQ ID NO : 271	146197403	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	NVLDWPKPQSDNDENSGNGRYGACCTEMDIV <sub>■</sub> ANS	198-230	202-214	222, 227
SEQ ID NO : 272	146197081	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	NVDDWPKPQDNDENSGDGKLGTCCS <sub>■</sub> MDIV <sub>■</sub> GNNA	197-229	201-213	221, 226
SEQ ID NO : 273	146197413	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	NVLDWPKPQSDNDENSGNGRYGACCTEMDIV <sub>■</sub> ANS	198-230	202-214	222, 227
SEQ ID NO : 274	146197309	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	NSDGWPKPQSDNDKNSGNGKYGSCCS <sub>■</sub> MDIV <sub>■</sub> ANS	196-228	200-212	220, 225
SEQ ID NO : 275	146197227	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	NSDGWPKPQKDNDKNSGNGKYGSCCS <sub>■</sub> MDIV <sub>■</sub> ANS	195-227	199-211	219, 224
SEQ ID NO : 276	146197253	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	NSEGWPQSDNDENAGTGYGTCCTEMDIV <sub>■</sub> ANS	200-232	204-216	224, 229
SEQ ID NO : 277	146197099	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	NVLDWPKPQSDNDENAGTGYGTCCTEMDIV <sub>■</sub> ANS	197-229	201-213	221, 226
SEQ ID NO : 278	146197409	uncultured symbiotic protist	NVLDWPKPQSDNDENSGNGMARGCCTEMDIV <sub>■</sub> ANS	198-230	202-214	222, 227

TABLE 4 -continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in active site loop residues in sequence identifier	Amino acid positions of active site loop residues in sequence identifier
SEQ ID NO: 279	146197315	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	NSDGWKPQSDNDKNSSENGKYGSCCSEN <u>M</u> DIV <u>R</u> ANS	196-228	200-212
SEQ ID NO: 280	146197411	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	NVLDWKWP <u>S</u> ND <u>D</u> EN <u>S</u> ENGRYGACC <u>T</u> ENDIV <u>R</u> ANS	198-230	202-214
SEQ ID NO: 281	146197161	uncultured symbiotic protist of <i>Hodotermopsis sjostedti</i>	NVQDNWKWP <u>S</u> ND <u>D</u> EN <u>S</u> ENG <u>G</u> YHGA <u>C</u> T <u>E</u> MDIV <u>R</u> ANK	201-233	205-217
SEQ ID NO: 282	146197323	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	NSDGWKPQSDNDKNSSENGKYGSCCSEN <u>M</u> DIV <u>R</u> ANS	196-228	200-212
SEQ ID NO: 283	146197077	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	NVLDWKWP <u>E</u> TD <u>E</u> NS <u>G</u> NGRYGTC <u>C</u> T <u>E</u> MDIV <u>R</u> ANS	201-233	205-217
SEQ ID NO: 284	146197089	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	NVEDWKWP <u>Q</u> DD <u>D</u> EN <u>S</u> ENG <u>K</u> LGTCCSEN <u>M</u> DIV <u>R</u> GNA	197-229	201-213
SEQ ID NO: 285	146197091	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	NVLDWKWP <u>S</u> ND <u>A</u> EN <u>G</u> TYGTC <u>C</u> T <u>E</u> MDIV <u>R</u> ANS	197-229	201-213
SEQ ID NO: 286	146197097	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	NVDDWKWP <u>Q</u> DD <u>D</u> EN <u>S</u> ENG <u>K</u> LGTCCSEN <u>M</u> DIV <u>R</u> GNA	197-229	201-213
SEQ ID NO: 287	146197095	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	NVDDWKWP <u>S</u> ND <u>D</u> EN <u>S</u> ENG <u>K</u> LGTCCSEN <u>M</u> DIV <u>R</u> GNA	197-229	201-213

TABLE 4 -continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of active site loop residues in sequence identifier	Amino acid positions of active site loop residues in sequence identifier
SEQ ID NO : 288 146197401		uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	NVLDWK <del>P</del> <u>Q</u> SND <u>E</u> NG <u>G</u> AC <u>C</u> T <u>M</u> DIV <u>R</u> ANS	198-230	202-214
SEQ ID NO : 289 146197225		uncultured symbiotic protist of <i>Neotermitess koshunensis</i>	NSDG <del>W</del> <u>K</u> P <u>Q</u> SND <u>K</u> N <u>S</u> NG <u>K</u> Y <u>G</u> SC <u>C</u> <u>E</u> M <u>D</u> IV <u>R</u> ANS	195-227	199-211
SEQ ID NO : 290 146197317		uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	NSDG <del>W</del> <u>K</u> P <u>Q</u> SND <u>K</u> N <u>S</u> NG <u>K</u> Y <u>G</u> SC <u>C</u> <u>E</u> M <u>D</u> IV <u>R</u> ANS	196-228	200-212
SEQ ID NO : 291 146197251		uncultured symbiotic protist of <i>Neotermitess koshunensis</i>	NSDG <del>W</del> <u>K</u> P <u>Q</u> SND <u>K</u> N <u>S</u> NG <u>K</u> Y <u>G</u> SC <u>C</u> <u>E</u> M <u>D</u> IV <u>R</u> ANS	195-227	199-211
SEQ ID NO : 292 146197319		uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	NSDG <del>W</del> <u>K</u> P <u>Q</u> SND <u>K</u> N <u>S</u> NG <u>K</u> Y <u>G</u> SC <u>C</u> <u>E</u> M <u>D</u> IV <u>R</u> ANS	196-228	200-212
SEQ ID NO : 293 146197071		uncultured symbiotic protist of <i>Reticulitermes speratus</i>	NILDWK <del>P</del> <u>S</u> ND <u>E</u> NA <u>G</u> AG <u>R</u> Y <u>G</u> T <u>C</u> C <u>T</u> <u>E</u> M <u>D</u> IV <u>R</u> ANS	200-232	204-216
SEQ ID NO : 294 146197075		uncultured symbiotic protist of <i>Reticulitermes speratus</i>	NVDDWK <del>P</del> <u>Q</u> D <u>D</u> <u>E</u> NG <u>G</u> AK <u>G</u> TC <u>C</u> <u>S</u> <u>E</u> M <u>D</u> IV <u>R</u> GNNA	197-229	201-213
SEQ ID NO : 295 146197159		uncultured symbiotic protist of <i>Hodotermopsis sjostedti</i>	NVKDW <del>K</del> <u>W</u> <u>P</u> <u>Q</u> E <u>T</u> <u>D</u> <u>E</u> NA <u>G</u> NG <u>H</u> Y <u>G</u> AC <u>C</u> T <u>E</u> M <u>D</u> IV <u>R</u> ANS	197-229	201-213
SEQ ID NO : 296 146197405		uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	NVLDWK <del>P</del> <u>Q</u> SND <u>E</u> NG <u>G</u> AC <u>C</u> T <u>E</u> M <u>D</u> IV <u>R</u> ANS	198-230	202-214

TABLE 4 -continued

Sequence Identifier (SEQ ID NO.)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in active site loop residues in sequence identifier	Amino acid positions of active site loop residues in sequence identifier	Position of catalytic residue in sequence identifier
SEQ ID NO : 297	146197327	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	NSDGWKPQDDNDENSGNGKYGSCCSEMDIV <del>E</del> ANS	201-233	205-217	225, 230
SEQ ID NO : 298	146197261	uncultured symbiotic protist of <i>Neotermitess koshunensis</i>	NSDGWKPQRDNDKNSGNGKYGSCCSEMDIV <del>E</del> ANS	195-227	199-211	219, 224

**117**

TABLE 5

Substitution(s)	Tolerance to 250 mg/L cellobiose % Activity in 4-MUL Assay (+/-Cellobiose) <sup>‡</sup>	Tolerance to cellobiose accumulation % Activity in Bagasse Assay (-/+BG) <sup>§</sup>	5
None	25%	60%	
R273K/R422K	95%	84%	
R273K/Y274Q/D281K/Y410H/ P411G/R422K	78%	ND	10

**118**

TABLE 6

Substitution(s)	Tolerance to 250 mg/L cellobiose % Activity in 4-MUL Assay (+/-Cellobiose) <sup>‡</sup>	Tolerance to cellobiose accumulation % Activity in Bagasse Assay (-/+BG) <sup>§</sup>	15
None	23%	74%	
R268K/R411K	92%	94%	
R268A/R411A	92%	95%	
R268A/R411K	97%	94%	
R268K/R411A	97%	102%	
R268K	ND	92%	20
R268A	ND	86%	
R411K	ND	89%	
R411A	ND	94%	

15

20

25

TABLE 7

SEQ ID NO.	Amino acid sequence
SEQ ID NO : 1	MSALNSFNNY KSAILIGSLL ATAGAQOQIT YTAETHPSL WSTCKSGGSC TTNSGAIITLD ANWRWVHGVN TSTMCYTGNT WNTAICD'TDA PNNKAGAODG ADYSGTYGIFT VTSNLSRFLN VEGWTESSN ANTGEIGNHGA CCAEFLDEWA NSISEALTHP PCDTPEL SVC QPSSKLGSVGS GAVINSDFCD AEISPTGETA SFSKTHGLAK MGAGMEAGMV LYMSLWDDYS VNMLWLDSTY PTNATGTVGA ARGSCPTSG DPKTYESQG SSYTFSDIR VGFNSTRSG GSSTGGSSTT TASGTTKA SSTSTSSTV GTGVAAHMGQ CGGQGWGTPT TCASGTTVY VNPYVSQCL
SEQ ID NO : 2	MYRKLAIVSA FLATARAQSA CILQSETHPP LTWKCGSSGG TCTQOTGSSV IDANWRWTHA TNSSNTCYDG NTWSSTLCPD NETCAKNCCCL DGAATASYG VTIISGNLSI GRYTOSQAKN VCARLYIMAS DTIQEFTLL GNESEVDYD SOLPCGNGA LYFVSNADAG GVSYKPTNTA, GAKYGTGYYC SQCPRDLPKI DLKPTAGQAN VEGWTESSN ANTGEIGNHGA CCAEFLDEWA NSISEALTHP PCDTPEL SVC QPSSKLGSVGS GAVINSDFCD GTCDDPGCDW NPYLGNLTKW YCPGSSWPL PTKKLTIVTQ RGTSGTAAIYR DIVEGNTPEQ PNAFLGSGYVG NEQELDUCYTA REAPFGSSR SDKGEGTQFK KATSGGMWLV MSUWDDYAN MLWLDSTYPT NETSTGPAY RGSCTSSVEV PAQVEOSSPN AKVTFNIKF GPICGSTGNDP GGNPGCGNPP GTTTRRAT TTGSSPQPTQ SHYGQCGGIG YSGPTVCAASG TTQCVLNYYY SQCL
SEQ ID NO : 3	MSALNSFNNY KSAILIGSLL ATAGAQOQIT YTAETHPSL WSTCKSGGSC TTNSGAIITLD ANWRWVHGVN TSTMCYTGNT WNSAICD'TDA SCAQDCAIDG ADYSGTYGIFT VTSNLSRFLN VEGWTESSN ANTGEIGNHGA CCAEFLDEWA NSISEALTHP PCDTPEL SVC TTDACGCTYS PNNZAGAQY VGYCDSOCPR DLKPTAGQAN VEGWTESSN ANTGEIGNHGA CCAEFLDEWA NSISEALTHP PCDTPEL SVC QPSSKLGSVGS GAVINSDFCD AEISPTGETA SFSKTHGLAK MGAGMEAGMV LYMSLWDDYS VNMLWLDSTY PTNATGTVGA ARGTCATSG DPKTYESQG SSYTFSDIR VGFNSTRSG GSSTGGSSTT TASRFTTSA SSTSTSSTV GTGVAHHMGQ CGGQGWGTPT TCASGTTVY VNPYVSQCL
SEQ ID NO : 4	ESACTLQSET HPLTWCQG SVVIDANWRW THATNSNTSC YDENTWSSL CDPNETAKN CCLDGAAYAS TYGTTISGNS LSDIDFTVQSA QXNGVQARLYL MASDTIOP TLUGNEFSD VDVSOLPGCL NGALYFVSM ADGGVSKYPT NTAGAKYGTG VCDQCPDL KFINQOANVE GWEFSNNAN SEMDIWQAN VSEALTHPC TTGQOEICEG DCGGFTSDN RYGGTCDPDG CDNPYRLEN TSPYGPSSF TLDTKLTIVTQ VTOFETSGAI NRYTQVNGVT PQPNMAELGS YSGNLLNDY CTAAEAEFGG SFSFDKGGLT QPKKATSGGM VLMWSLWDDY YANMLWLDST YPTNETSSTP GAVRGCSCTS SGVPQVESQ SPNAKUTFSN IKPGPIGSTG NPSG
SEQ ID NO : 5	MASSPOLYKA LLPPSSLLSA VOAQKVQNO AEVHPGSHTWQ TCTSSGSCFT VNGEVTDIAN WEWLHTVNGY TNCYTGNEWD TSICTSNEYC AEQCAVDGAN YASTYGITIS GSSLRLNFVQ QSQRNIGSR VYLMDDDTY TMFLYLNKEF TEDVDSLELP CGLNGAVYFV SMDADGKSR YATNDAGAKY GTGYCDSOCPR RDLFKINGA IVEGEVNHG SCCAEMDIE ANSISTAFTP HPCDPGOTL CTGSGCGGTY SNDRYGGCD PDGCDENSTR PDGCDENSTR VTDTINPSN VTNQPLTEK REYVFGVVI PNISITTEFC ESOKELFGDV DVRAHGENWA CNGAALFQGM VLVLSWDDN YSNMLWLDN YPTDADTOP GIAORGTCPTD SGVPSTEAEQ YPNAYVYEN IKFGPIGSTF GNGGSGPTT TTTTSTATST TSSATSTAG QAHWHEQCGG NGWGTGTVCA SWPACTVNS WYSQCL
SEQ ID NO : 6	MYRALATASA LIAAVRAQY CSLTQSKPS LNWSKCTSG CSVYKGSVTI DANWRWTHQV KWDTSYCTSG KVCAEKCCLD GADYASTYGI TSSDQLSLS FVTKGPTSTN IGSRTYIMED ENTYOMFOLI GNIEFTDYDVA SNIGCNGA LYFVSNADAG GKAYKPGENA GAKYGTGYYC AQCPRDVKFI NQANSDWQ PSDSDVNGI GNLGTCPEM DIWEANSLST AVTHPCITKL TQHSTGDCS GGTISNDRYG GTCDAADGCDP NSTROGNKTF YCPGSGFENV TTKKVTVTO FHKGSGNRILS EITRLYVQNG KVIANPSKII AGVPNSLTA DRCTKOKKV NDPDDFTKQ AWSEMDSALE APWLYMISLW DSTYPTDSTK LGSQRGSCST SSGVPADEK NVPNSKVAFS NIKRPIGST YKDGTPTN PTNPSEPSNT ANPNPGTVQ WGQCGGNSY CPTACKSGFT CKKINDFYSQ CQ
SEQ ID NO : 7	MLASTFSYRM YKTAIIAAL LGSGQAQVQG TSQAEVHPSM TWOSETAGGS CTTNGKUVI DANWRWVHKV GDYTMCYTGNT TWDDTICPDD ATCA,NCALAE GANTESTYGV TATGNSLRLN FVTTSOONI GSRLYTMKDD STYEMFLKLN QBTFTDV S NLPQCGINGAL YFVANDADGG MSKYPTNKAG AKYGTGYYC QCPRDLEKIN GOANTEGWQ P SNDANAGTG NHGSCCAEMD IWEANSISTA FTHPHCDTPG QMCTGDAGC GTYSSDRYGG TCDPDDGCDN SPROQNTFYF GPGMTVDTKS KFTVTVQFIT DGTGSSGTLK ELKRFYQONG KVPSBESTW TGVSNSNITT EYCTAQKSLF QDNVVEEHG GLEGMGALA QGMVLYMISLW DDHSANMLWL DSNTYPTASS TTPGVARGTC DISSGVPADEV EANHPDAYVV YSNIKVGPIG STPNSSGEPN GGGTTTTAG NPGGTVGAQH YGQCGGGIAT GPTTCASPYT CQKLNDYYSQ CL

TABLE 7-continued

SEQ ID NO.	Amino acid sequence
SEQ ID NO : 8	M LPSTISYRI YKNAFFFAAL FGAQAOAQKVG TSKAEVHPM SM AWOTCAAGT CITTKNGKVVII DANWRWHDV KGYTCYGTN TMIAELCPDN ESCAENCAL GADTAATIGA TISGNLSSK FVTQSQRN GSRLIIMKD NIETFVKLN QFTFDVDS NLPGLINGG YFVNNDADGG LSRYTGNEAG AKTGTGYCD S CPCRDLKFN GLANTBEGWT PSSSDANAGNG GHSSCCARND IWEANSTSTA YTPHPCDTPG QAMCNGDSCG GTYSSDRYGG TCDPDPGDN STRQGNKSYF GPGMTVDTK KMIVTUTQFLT NDGTATGUL EIKRFTVQDG KVIANSESTW PNLEGENSLTN DECKRQKTVF GDMDTFSKHG CMEGMGAALA EGMLVLYMSL DDHNSNMMLW DSNSPTGTCS TIPGVARGSC DISSGPDKL EAHPDASTV YSNJKVGPIG STENSGGSNP GGSITTKPA TSTTITKATT GTGVAQPWA CGGIGYSGP QCAAPYCTK QNDYSSQCL
SEQ ID NO : 9	MHPSLQTILL SALPTTAHQ QACSSKKPHTH PPLSWSRCSR SGCREVOGAV TVDANWLMNT VDGSQNCTG NRWDTSICSS EKTCSSECCI DGADYAGTYG VTTIGDAISL KEVQOGYISK NVGSRLYMK DESRYEMFTL LGEHTEFYDQ VSKLGCGUNG ALYFYSMD ED GMKRFEPMNK AGAKRGTGYC DSCQCPDVKP IPSKSDANAG IGSLGACRE MDIWEANNA SAFTPHICKN SAYHSCTGDG CGGTYSKNRV SGDCDPDGCD ENSYRLGHTT FYPGPGRKFTI DTTRKLSVTQ OFLIGRDSL REIKRFTYON KGVIPNVSVR VRGPENSIT OGRCNAQKEM FGAHBSFNK GGMIKGMSAAV SKPMVLVNL WDDHNSNMMLW LDSTYPTNSR QRSKSKRGSP SSAPGD-TVVFV SNIKRGPIGS TFRGK
SEQ ID NO : 10	ESACTLQSTT HPPLTWOKRS SGGTCTOOG SVVIDANWRW THATNSSTNC YDENTWSTL CDPDNECAKN CCLDGGAYAS TYGTTSGNS ISIDEVQSOA OKNVGARLYL MASDFTYQEFL TLLGNEFSD VDVSQPLCGI NGALYFVSM NTAGKXKPT NTAGKXKPT YCDQCPDRD KFINQGANE GWFRSSNNAN TGIGGHGSC SEMDIWQANS ISEALTHPHC TTYQGEICRG DCGGTYTSND RYGGTCDPDG CDWMPYRLGN TSPYGPQSSP TLDTTKLTV VTORETSGAI NRYYQVQNGVT FQPNAAELS YSGNELLNDY CTAAEAEFGG SSESDKGGLT QPKTATSGGM VLVMWLDDY YANMILWLDST YPTNETSSTP GAVRGCSCS SGVPQVOESQ SPNAKVTFSN IKFGP1GSTG NPSCG
SEQ ID NO : 11	MHORAILFSA LAVAANAOQV GTQKPETHPP LTWKQCTAG SCSQOSGSVY IDANWRLHS TKTDTINCYTG NTWNTTELCPD NESSAQNCAV DGADYAGTYG VTISGSELKL SPVTGANVGS RLMLQDDET YOHNELNNE FTDPDVDSNL PGGLNGALYF VAMDADGMS KYPSPNKAAGK YGTGTCDNSQ PRDJKIFNGM ANVEGNWPS NDKNQAGVGH GSCSCCEMDI EANISSTAVT PHPCDMSVND VSGCSLSSDF CTAKKAPGD DPDGCDENP RMGNESFPC GKVUDTSEPS TVUTQPTAD CTGDSMSEI KRLYQNSQV TANISVSDV YSATRHYAGC EDIFAKHGGF SGMGKALSEM VLIMSINVDDH HSSMMWLDST YPTDADSPK GVARTCHHG ACDPEKVEQS HDASVTSFN IKFGPIGSTY KA
SEQ ID NO : 12	MYRSJLFATS LLSLAKGOLV GNLYCKGSCT LNWSKCTSG CTNVAQSITY DANWRWTHIT SGSTNCSGN EMDTSLCSTN TDCATKCCVD ROLLGETHH QGLYSTNCG RTYLQDSTD YQLKRETSQ EFTFDVDSN LPGCLNGALY FVSDMADGGL KKYPTNKGAG KYGTYCDAQ CPRDLKFING EGVNGWPOE KNDQNAQGYG HGSCSCCEMDI WEANSVSTAV TPHSCSTIEO SRCDGDEGGG TYSADYAVG CDPOGCDENS YRMGYKDFYK KGKTVTDTSKK FTVTQFIGS GDAMELKRYF VONGKTIQP DSTIPGVTON SITTFPCDAQ KKAFGDKYTF KDKGGMANNP LWLDSTYPTD KNPDTDASSG RGECAITSVG PADVEQHQD ASVYNSIKE GPINTTFG
SEQ ID NO : 13	MLAKPAAIA LVAZANAQAV CSLTAETHPS LNWSKCTSG CTNVAQSITY DANWRWTHIT SGSTNCSGN EMDTSLCSTN TDCATKCCVD GAEKSYSTYGI QTSGNSLSQL FVTKGSTSKN IGSRTYLMG ADAYQGFLL GNEFTFDVDSV SGTCGLNGA LYFVSMIDLG GKAYTKNNKA GAKYKTYCDD AOCPRDLKVI NGIANVSGWT PSTNDANGI GDHGTCCESEN DIMEANKVST AETPHPCITI EOHMRCGDSC GGTGDDRGZ GTCTCADCDFD INSTRNGTIVT YGEGKTVTDTSKK FTVTQFIGS GDAMELKRYF VONGKTIQP DSTIPGVTON SITTFPCDAQ KKAFGDKYTF KDKGGMANNP DDFENKKGGL VLYLMSIWLD HAANMULWLS AKPSSTSSTS NPSGTTAAHW AQCGGIGFSQ PTTQSPYTC QKINDYYSQC V
SEQ ID NO : 14	MFKKVALTAI CFIIVQAQQ VGREVAENHP RLPWQRCTRN GGCGTIVSNGQ VVLDANWML HYTDGYINCY TGNSNISTVC SDPTICAQRCN ALLEGANYOOT YGTTTNGDAL TIKFLTTSQ TNVGAERYVLM ENERYOMFN LINKEFTPDV DVSKVPGIN GALYFTOMDA DGMISKOPNN RAGAKYGTGY CDSQCPDJKF FIDGVANSAD WTPSEDPNA GRGRYICCA EMWIWEANSI SNAYPHPCR TQNDGYZQRC EGRDNQPRY EGLCDPDPGCD YNPERMGNKD FTYGPCKTVDT NRKMTVUTOF ITHDNTDTGT LVDIRRYYO DGRVIAUPT NFPGMPAHD SITEQFCTDQ KNUFGDYSSF ARDGGLAHMG RSIAKGVHLA LS1WNDHGAH MLLWLDNSNPT DADPNKPGIA RGTCPTTGGT PRETEQNHDP AOVIFSNIKP GD1G3TSFGY
SEQ ID NO : 15	MYSAAVLAPE SFLIGGAGQO YGTSTAAETHP ALTVQKAG GTCIDBESDI VLDANWML STSGSTNCTY GNTWDITLCP DAATCTTNA LDGADYEGTY GITTSGDSLK LSFVTSNSV SRTYLMSET TYKEFALLEN EFFTFTVDSK LPGCLNGALY FVPMADGGM SKYPTNKAGA KGKTYCDDAQ CQDMKFEYG TANVEGHPD SNSANSGTGK IGSCCSEFD WEANMSQD TPHVCTVDSQ TACTGDDCAS NTGVDGDGC DENPYRMGTQ TFYGSGMTID TSKEPSVUTQ FITDDGCTGT TLTEIKRFVY QDDVYEQPS SDISGVSNS ITDDETAQK TARSDTDYFT QNGGNAAMGK KMADGMVLYL SWDDTNVN LMWLDSDYPTT KDASTPGVSR GSCATDSGVP ATVEAAGSA YTFTSIIKG PIGSTENAPA

TABLE 7-continued

SEQ ID NO.	Amino acid sequence
SEQ ID NO : 16	DSSESVSASS SPADIASSS SASIAPVSS VAAIVSSAQ AISSRAPPV SSEAQISSAA PVVSSVVSSA APVATSSKS KC5IVSSTLK TSVAAPATSA TAAVAVATSS AASSTGSPV SEGTCTGKTC YGNCTGKPL YSSQCVASS
SEQ ID NO : 17	MWFORCCTG GS SCTTNVNGE IVTIDANWRWI HATGGYTNCF DGNEWNTAC PSENACTKNC AIEGSDYRGT YGITSGNSL TLKFITKGQY SNTVNSRTYL MDTNNYEMF NLIGNEFTFD VLDSOLPCGL NGALYFVSM P EKGQTGAK YGTGKLSQCS VHI SKTLTDA CARDLKFVGG EANAGWQAS TSDENAGYKG KGACCAENDW WEANSNSTAL TPHECOEGY AVCEESNCGG TYSLDIYAGT CDANGDFFP YRVSNKDFYFG KGKTVDTSKK MTYVTFQFLGT GSIDLTELKPV YDQGRKVN PEPTGAVTAN KQEVEEVYI QPAVGMAS MGKMAQGMV LVMSLWDDH SNMILWLDSTY PTDRDPEPSG AARGEFAVITS GAPARVEANV PDAVSMFNI KRGPIGSTEQ QPA
SEQ ID NO : 18	MQIKSYIQLV AAALPLLSSV AAOQAGTTA ENHPRMTWK CSGPNCOTV OGEMVIDANW RWLHNNGONC YEGNKNTSQC SSATDCAQRC ALDGANYQST YGASTSGDSL LKPVTKHNEY GTNQYKQMF NLIGNEFTFD VLDSKVEBGI NSALVBCPMDA DGGISRYPGN MRAGAKYGTG YCDAQCARDL KFPGKANNE WTGSTDNA GAGYGTCCS ENDIWEAIDM ATAFTHPCT LIGQSRCEGD SCGGTYNSER DRFAGYCDAN GCDTNPYNGK VDTNRAKPTVV SRFEENRLSQF FVQDGKRLC VBPPTWGLP NSADTPELC DAQPRFDDR NRPAETGGFD ALNEALTIPM VLVMSIWDDH HSMLWLDSY YPEKASHJPG GDRGPCEPTS GPVPEAEVQY PMQVQVWENI RFGPIGSTVNV
SEQ ID NO : 19	MRTAKPATLA ALYASAAAQO ACSLTTERHP SLWKKCTAG GOCTVQASI TLDNNWKRTH QVSGSTNCYT GNKMDTSICL DAKSCAQNC VDGAQYTSY GIFTNGDSU LKPVTKCQYS TNVGSSTYLM DGEDKTYQTFE LIGNEFTFDV DVSNGICGLN GALVYFMSMDA DGGISRYPGN KAGAKYGTG CDACQCPDKI FINGEANIEG WTGSTDNA GAGYGTCCS ENDIWEAIDM ATAFTHPCT LIGQSRCEGD SCGGTYNSER YAGVCDPDGC DFNSYRQEN VTIKIIWVTT FLKDANGDLG EIRGFYVQDG KLIPIHSNITQ DWCRQKVAF GDIDDEFNRKGM QMOMGKALA GPNVLYMSNWL DSTEFPVDAAG KPGAEGRACP TTSGVPAEEV AEAPNSNVF SNIREFGPIGS TVAGLPGAGN GGNGGNPPP PTTTSSAAPA TTTTASAGPK AGRWQOQGI GFTGPTQCE PYTCTKLNDW YSQCL
SEQ ID NO : 20	MQIKOYQYL AAALPLVNA AAOAGTQTO EHTPRULSWK CSGGNCOTV NAEIVIDANW RWLHNDSYON CYDGENWTSA CSSATDCAQK CYLEGANYLQ TYCUTSYTQKA LTLKPFVTKHNEY IMLNSDQKYM FLTMNEDQKYM LNSALVYFVAM BEDGMRSTIS SNKAGAKYGTG GYCAQCARDL LKPVFGKANI EGWRPSTNDIA NAGVYGPYAC CAEIDWNSN AYAFATPHG CLNMYHVC E TSNGGTSE DRFEGGLCDAN GCDTNPYNGK VDTNRAKPTVV TRFEENKLQF FFIQDGKRLC I PPPTWGLP NSSATPELC TNLSKVEDDR DRYEETGGFR TINEALRTPM VLVMSIWDDH YANMLWLDSV YPPERAGQGP AERGPACAPS GVPAYEBAQF PNAQYIWNSI RFGPIGSTVQ V TVPGLDGSTP SNPATVAPP TTTTSSVRS TTQISIPTSQ PGGTTQKNG QCGGIGTGC TNCVAGITCT ELNPVYQCL
SEQ ID NO : 21	MMYKKPAALA ALVAGAAAQO ACSLTTEHPL RLTWKRKTSG GNCSVTNGAV TIDANWRWTH TVSGSTNCYT GNEWDTSCS DGKSCAQTC VGDADYSTSY GITSGDSLW LKPVTKHNEY TNVGSRVLM ENDKTYQMF LIGNEFTFDV DVSNGICGLN GALVYFMSMDA DGGMSKYSQN KAGAKYGTG CDACQCPDKI FINGEANIEG WTGSTDNA GFGRTGSCCS ENDIWEAIDM ATAFTHPCT LIGQSRCEGD SCGGTYSSER YAGVCDPDGC DFNSYRQEDK TYKGMNTWD TPKMVTWTO PHRNISAGVYK TFKMVTWTO PANPNSITQ EWCAQKVAF GDIDDEFNRKG GMQAMSKALE GPNVLYMSNW DDHYANMLW DSTPVIDKAG TPGAEGRACP TTSGVPAEEV AEAPNSNVF SNIREFGPIGS TVPGLDGSTP SNPATVAPP TTTTSSVRS TTQISIPTSQ PGGTTQKNG QCGGIGTGC TNCVAGITCT ELNPVYQCL
SEQ ID NO : 22	MYRNFLYAA LLSTYARSQV GTOTTETHPG MTWQSCTAG S CTCSDSDNA CASNCVADGA DYKGTGIGTA SGNSLQKFI TKGSYSTNIG SRTYLMASDT AYOMKFKDN KEFTDVLDS GLPGCGNGL YFVNMDDEGG LKCKYSGKAG AKYGTGTCDA QCPRLKIFIN GEHNGEWGP SDNDANAGVG GHGSCCAEMD IWEANSLSTA VTPHACSTIE QTRCDGDCG GTYSADRYVQ VCDPDGDFN AYRMGKYNFY GKGMDVDTSK KFTVYTOFIG TGDAELEKRF YYQGGKTEQ PASTIPGEV NSITTKFCDO QRQVFGDRYI YKEKGFTANN AKALQGMVL VMSLWDDHTS NMWLWLDSTYP TDKNPDTDLG SGRGSDVKS GAPADEVKS PDAVYTSNI RFGPINSTY
SEQ ID NO : 23	MLGKTAIASL SFLAJAKGQQ VCREVAENDH RLPWQECTRN GGQTVNSQ VVLDANWHL HVTDGYTNCY TGNSNNSVC SDGTTCAQRC ALEGANYOQT YGTTTSGNSL TMKFLTISQG TNVGGERYL MENEYQOMM LIJKEFTEFDV DVSKVPGIN GALVYFOMDA DGMNSQOPIN RAGAKYGTGY CDSCQCPDKI FIDGVANNSVG WEPSETDNA GRGRYGICCA EMDIWEANSI SNAUTPHPCR TQNDGQYQRC EGRDQNQPRY EGLCUPDGCQ YNPERMGKID FVGPQKLTW NRKMVTWTO ITHDNTDTGT LWDIIRLYVQ DGRVIAUPT NFGPMAPHD SITEQCTDQ RNLFQDYSSE ARDGLGHLNG RS LAKGHLVA LS WINDHGHAH MLWLDNSNPT DADENPKGIA RGTCPITGGT PRETEONHPD AQVIFSNIKE GDIGSTESGY

TABLE 7-continued

SEQ ID NO.	Amino acid sequence
	DGCDENSFRM GDKSFYGPQM TVDTNQPIV VTQFIDNGS DNGNLQEIRR IYVONGQVIQ NSNVNVLPGID SGNSISAEPF DQAKEARFGE RSFOURGGLS GMGSALDRDM VLVLSLNDH AVNMMLWDSL YPLDASPQSP GIERGTCSR SKRPEDVAN AGGVQVVSN IKPFDINSTF NNNGGGGNP SPITTRNPSP AQTMWGQQG QWITGPACQ SPSTCHVIND FYSQCF
SEQ ID NO : 24	MYRNALASL SLFAARAQO AGTVTTETH SLSWKTCGT GGTSCTTAG KITLDANRHW THVTGTYTNC YDGMSWNTTA CPDAGATCTRN CAVDGADYSG TYGTTTSENS LSTIKFVTKGS NSANIGSRVY LMESDITYKOM FNLIQOEFTF DVDSVKLPCG LNGALYEVEM AADSGIKGK MRAGAKYGTG YCDSQCPRDVK GMPSQADPACM AGSGKIGLST PENDWEAVS I STAYPKHPC KGTFQBCD DVSCDGSNR YSGHCDKDGC DENSYRMVTK DEYGPQATL DTTKKMTVYTQ FLSGSSSTLSE IKFYVONGK VKTNSDAIE GVTYNSITES FCQAQTARG DTNSFKTLGG LNEMGASLAR GHVLYNMSLWD DHAVNMLWLD STYPTNSTKL GAQRGTCID SGKPEDEVRN HDATVVFSD I KFGP/GSTF QQPS
SEQ ID NO : 25	MVDIQLATFL LLGVGVQAAQ QVGYTIIPHN PLLATOSCTA SGGCTTSSK IVLDANRHWI HSTLGTSCL TANGDPTLC PDGJTCANC ALDGEYSYSSL YGTTTSGSAI RLUQFVCTNII GSRLVFLMADD THYRTFQLIN QELAFDWDY KLPCPGHNGAL XEVANDADGG KSKYTPGNRAG AKYGYGYCDS QCPRDVQFVIN GOANVQWNA TSATGTSY GSCCTFELDIW EANSNAALIT PHCTINNAQT RCSGNCNTSN TGFDADGCD ENSPFLGNTT FLGAGMSYDT TKTFTVYTOF ITSNDNTSTGN LTETFRFYQ NGVNIPIVSVV NYTGIGAVNS ITDPFESQOK KAFIETNYPA QHGGLAQLQO ALRGMVLAFL STSSDPANH LMWLDNNEPPS ANPAVGVAR GMCSITSGRP ADVGILNPSP YVSFLNFKFG SIGHTFRPA
SEQ ID NO : 26	MHQRALFLSA LAYAANAOQV GTQTPETHP LTWQKCTAAG SCQSOSGSTVY IDANWRHWHS TKDTNTCYTG NTWNTELPD NESCQNCAL DGADYTAGTYG VTISSGSEKL SPVTFGANVGS RLYLMQDDET YQHENLHNHE FTPDVDSNL PCGLNGALYF VAMDAGGMS KYPNKAAGK YGTGFCDSQ PRDIFINGM GSKCPEMDIW EANSISTAVK TLYVNGVAVS KTDGSLASEI KRLYVNGVQV TANVSENVAG VSGNSITSDF CTAKQKAFGD ED1PAKHGGL SGMGKALSEM DPDGCDENPF RMGENESYFP GP KRVTDTSKVM TVVTOFITAID YPTDADPSKP GVARTGQH AGDPMENVEQ HPDASVTSN IKFGP/IGSTY EG VLMISWDDH HSNNMLWLD STYPTDADPSKP GVARTGQH AGDPMENVEQ HPDASVTSN IKFGP/IGSTY EG
SEQ ID NO : 27	MFRATLIAF TMAAMVFGQO VGTNTAENHR SLTISQCTKSI VLDANWRHWLH STSGYNTCYT GNQMDATLCP DGKTCAAANCA LDGADYGTG GITASGSSN HYQNLINQ EFFDUMSNL ITPCGHNGALY ISAMADGCM AKYTPNKA RYGTGCDTSQ CPTLKIFING SANAGHNTYG TCCTENDMIVE ANDDAATT HPCTTNAOTR CGSDCTRTD GLCDAADGDF NISPRNGDQTF IKGCLTVDTS KPTVUTQEI TNDGTSAGTL TEIRRLYVQN GKVNQNSVK I PGIDPNTSI TDNECSQKTI ARGDTNFAQ HGGLKQVGEA LRTONVLAIS IMDYAANML WLDNSNTPK DPSTPGVARG TCATTGVPAA QLEAQSPNAY VFVSNIKFGD LNTTYTGTVS SSSVSSSSS TTSSSHSSS STPTQPTGQ TVPQNGQCGG IGYTGSTTCI SPITCHLN P YSQCY
SEQ ID NO : 28	MQIKSYIQL AAQAGQTIA ENHPRMTWKR CSGP/GNCQTY QEGNVIDAWN RWLHNHNGC YEGNWKNTSQC SSATDCAQRC ALDGANYOST YGASTGSDSL TLKFVTKHEY GTNIGSERFYI MANONKYMPE TLUNNEFADP VDLSKYECGI NSALYEVAME EDGGMASVPS GAKYGTG/YCD SQCPRDVKI NGOANVBEW GMPSQADPACM AGSGKIGLST PENDWEAVS I YAYAFTPHAC GSKNRYHICE TNNGCTYSD GTCDDPDGDF NPYOPGNSF YCPGKIVDTS SKFTVYTOF TDDGTPGSTL TDKRFTYTON GRVYIPOSEST ISGVTCNTSIT TEYTAQKA FGDNTGFFTH GGLKLTISQAL AGCMVLMWSL WDDHAANMLW LDSTYPTDAD PDPFGVART CPTTSGPAD VESQNENSTV LYSNIKVGP1 NSTFTAN NRAETGGFD ALNEALTIPW VLMSLNDH HSNNMLWLD SSPTQPTGQ TVPQNGQCGG IGYTGSTTCI SPITCHLN P YSQCY V
SEQ ID NO : 29	MYRKLAIVSA FLATARAGSA CPTLQSETHP LTWQKCSSGG TCTQOTGSSVY IDANWRHWTH TNSSTINYDG NTWSSTLCPD NETAKRNCL DGAAYASTYG VTISSGNSLII GRYTOSQKN VGARIYLMAS DTYQEBTLL GNEFSFVDVY SOLPCGNGA LYFVNSNDAG GVSKYPTNTA MRAGAKYGTG YCDAQCARDLI KFJGGKANE GWRPSTNDP AGVGMGACC AEIDWWEINA YAYAFTPHAC GSKNRYHICE TNNGCTYSD DRFAGYCDAN GCDYNPYNGK NEDFYGKGT VDTNREKTVV SRFEERNLSSO FFYDGRKLE VPPPTVGLP NSADTPELC DAQRFVDDR NRAETGGFD
SEQ ID NO : 30	MYRKLAIVSA FLATARAGSA CPTLQSETHP LTWQKCSSGG TCTQOTGSSVY IDANWRHWTH TNSSTINYDG NTWSSTLCPD NETAKRNCL DGAAYASTYG VTISSGNSLII GRYTOSQKN VGARIYLMAS DTYQEBTLL GNEFSFVDVY SOLPCGNGA LYFVNSNDAG GVSKYPTNTA MRAGAKYGTG YCDAQCARDLI KFJGGKANE GWRPSTNDP AGVGMGACC AEIDWWEINA YAYAFTPHAC GSKNRYHICE TNNGCTYSD DRFAGYCDAN GCDYNPYNGK NEDFYGKGT VDTNREKTVV SRFEERNLSSO FFYDGRKLE VPPPTVGLP NSADTPELC DAQRFVDDR NRAETGGFD
SEQ ID NO : 31	MYQRALLFSP FLAARAQOA GTVTAENHPS LTWQQCSSGG SCCTQNGKVV IDANWRWHT TSGYNTCYTG NTWDTSICPD DVTCAQNCAL DGADYTAGTYG VTISSGNSLII GRYTOSQKN VGARIYLMAS DTYQEBTLL GNEFSFVDVY SOLPCGNGA LYFVNSNDAG GVSKYPTNTA GAKYGTG/YCD SQCPRDVKI NGOANVBEW GMPSQADPACM AGSGKIGLST PENDWEAVS I YAYAFTPHAC GSKNRYHICE TNNGCTYSD GTCDDPDGDF NPYOPGNSF YCPGSSPFLD TTKKLTLVYTQ FETSGAUNRY YVONGVATEQO PNAELGYSVG NGLNDUYCTA EEAPEGSSP SDKGSLTQFK KATSGGMWLV MSWLDSTYV NETSTPQAV RGSTSTSSSEV PAQVEOSPN ARVTSNIKF GP1GSTGDBS GGNPFGCNPQ PP GTTTRRAT TTGSSPQPTQ SHYQGCGGIG YSGPTVCAAG TTCQVLLNYY SQCL

TABLE 7-continued

SEQ ID NO.	Amino acid sequence
SEQ ID NO : 32	GTCDPDGCF NPYFQGNISF YGPQGIYDTS SKPTVVTQFI TDDGTPSGTL TEIKRFTYQN GKVIPQEST ISGVITNSIT TEYCTAQKAA FGDNTRGFFTH GGLQKISQAL AQGMVILVMSL WDDHAANMLW LDSTYPTDAD PDTDGYTARGT CPTISGVAD VESQYPSVY IYSNIKVGP1 NSTFVAN MIRKTTLAA LVGVYRGQAA CSLTAETHPS LTWOKCGSSG SCNTVAGSYT IDANWRWHTH TSGYTNCYTG NWKDTISCT NADASKCCV DGANYQQTG ASTSGNQSLA QVTQSSQRN VGSRLYLLES ENKYOMNLN GNEFTFDYDA SKLGCLINGA VYFVSMADG QSKYSGNKA GAKYGTGTYC SOCPRDLKYI NGAANVYSGW PSSGDANSV GNMGSCCAEM DIWEANSTIST AYTPHPCSNN AOHSKRGDC GGTTSVRYA GDCCDPGCDF NSYFQGNRF DDHKVTVLW TD STYPTTAP GAARGSCSTS SKRPSVQSO TGVATYVSN IKRPPIGSTF KSS DQNVYTNQRGG LRONGDAIAK GMVLYMWSYD DHHSQMLW SDYPAQMLW SDYPTDADPT TPJTGARTCIP TDSGVPDSVE SQSPSIVTY SNIKFGPTNS TFTAS
SEQ ID NO : 33	MLRRAILLSS SAILAVKAAQ AGTATAENHP PLTWOECPAT GSCTTONGAV VLDANWRWHTH DVNGYTNCYT GNTMDPTYC P DDETCQAONCA LDGAYEGTY GVTSSGSLKG LNFVFTGNSV SNNANTGIDG RLYNMADD EFTEDVDSN LPCGLINAGV FVPMADGGV SKYPRINKAGA KYGTGTYCDSQ CPDRKFDG EANVEGHOQS SNNAUTGIDG HGSCCAEMDV WEANSTISTP THPCDTPGQ TMCSGDICGG SKYPRINKAGA CDPDGDFNPK YRMNTSFVG PGKLIDTQP FTVVTOPLD DGDGTGTLW IKEYIQNSN VLPQPNSDIS GVTGNSITTE FCTQKAFAG DTDDFSQHGG LAKNGAAMQ GMVLYMWSYD DYAAQMLW SDYPAQMLW SDYPTDADPT TPJTGARTCIP TDSGVPDSVE SQSPSIVTY SNIKFGPTNS TFTAS
SEQ ID NO : 34	MHORALLFSA FWTAQVQAA GLTAETHPS LTWOKCGSSG TCBEOKGSSV LDSENWRWHTH VDGSTINCYT NTWDATLCPD NECASINCAI DGDATEGTY VVTSGDAITL QPVTKGANIGS RLYNMADD SYQPENLHN EFTEDVDSN LPCGLINAGV AKYTSNKAGA KYGTGTYCDSQ CPDRKFDG EANVEGHOQS SNNAUTGIDG HGSCCAEMDV WEANSTISTP THPCDTPGQ TMCSGDICGG YSSERYAGTC DPDGCDENAY RMGNEFSPV SKLVDSSSPV QMVNGALSM VLTLSINDH HSSMWMLDS YPEDADATP GVAEYCEPH AGDPKEVESQ SGSATVTSN IKYGPIGSTF DAPA DDIIFQHGGL QMGNAALSM VLTLSINDH HSSMWMLDS YPEDADATP GVAEYCEPH AGDPKEVESQ SGSATVTSN IKYGPIGSTF DAPA
SEQ ID NO : 35	MASLRSKTY KNALLAIAFL GAAQAOQVGT STAEVHPSLT WOKTAGGSC TSQSOKVYID SNARWVHTNG GYTNCYTGND WDRTLCPDDV TCATINCALDG ADYKGTGTYC PCRDLKING QANVEWNGS SSDVNGTEN KYNFMOLINQ EFTEDVDSN LPCGLINAGV FVAMDDEDGM ARYPNTKAGA KYGTGTYCDAQ CDPGCDENAY RMGNEFSPV YRMQNOFYF PSKIVDTEP FTVVTOFITH DGTSTGTLSE IKRFYVONGK VPOVSTIS AVTSNTIDS TYSSTDRYGGT DPDGCDENAY RMGNEFSPV YRMQNOFYF PSKIVDTEP FTVVTOFITH DGTSTGTLSE IKRFYVONGK VPOVSTIS AVTSNTIDS ECSSAQKTFK DTVPFAKGG MAGMAGLAE GMVLVMSLWD DHAANMLW STPTSAST TPGAAEBCD ISSGPBSDVE ANHSAVYVYI SNIKVGPLGS TFGSTDSSGG TTTKVTIT ATTKTITTTGP STTGAAHYAQ CGGQNWQTPT TCAСПYTCOR QGDYTSQCL
SEQ ID NO : 36	MVSAKPAALA ALVASASACQ VCSLTPEBSPH PLTWOCSAG GSCTNIVAGSV TLDSENWRWHTH TLQGSTINCYT GNEWDTISCT TGKCAONCC VEGAIIATY GITISGNQLN LKPVTEGKSY TNVGSRTLM ENATYQGFLN LIGNEPTD VDSNIGGLN GALYFVMSMDL DGGIAKYSGN KAGAKYGTGY CDACCPDOK FINGEANLGE WNPSTDVNA GAGRTGTCCS ENDIWEANIM AVATYHPSCT ILDORCEGE SCGETYSSDR YGGVCDPDG DFNSYRMKNG EFKYKGKTVT TIKKMTVTO FLKRNAGLSE FIRKFWYNG VVITPSVSI PGVPMONST QDWDAQKIA FGDPDNTAK GGURQMGIAL DCPMVLMSI WNDHAAMHLW LDSTPVDAA GRGAERGAC PTTSGPSEV EAEPNSVNA FSNIKFGPIG STPNSGSTNP NPISSSATT PTSIYVSSTS TAAQOPTSAP GTTVPRNGC GGQGYTGTQ CVAPYTCVVS NQWYSQL
SEQ ID NO : 37	MFPYTALVFSP SFLSVVLAQO VGTLLTAETHP QLTVOQCTRG GSCTTQQRSY VLDANWRWHTH STSGSNMNCYT GNTWDTSCLP DAATSRNCA LDGADYSGTY GITISGNALT LKFVTHGSPY TNIGSRYLLI ADDSHYOMEN LKKEPFTDYD DVSLQPGLN GALYFVMSMDL DGGIGRPUN KAGAKYGTGY CDSCOPDHDLK FINGEANLGE WNPSTDVNA GAGRTGTCCS ENDIWEANIM ASAYTHPSCT VTPPRCQGN DCGDGDNRYG GVCDKDGCDF NSFRNGDKNE LGPKTNTN SKFTVVTQFL TSDNTTSGLT SEIRLTYON GRVIONSKVN IGMSTLDS ITEFCSTOK TVFGDINSFA SKGGRLANGN AFDRGMWLV SIWDDHEAR LWLDNSNPLD KSASAPGVAR GTCATSGEP RDVEOSDNA QVIFNSNIKG DIGSTYNS
SEQ ID NO : 38	MYRALATASA LIAAVRAQGY CSLTQESKPS LNWSKCTS SGSYKGSVTI DANWRWHTH SGSTNCYTGN KMDTSTVCTSG KVCAEBRCCLD GADYASTYGI TSSGDSLHS FVTKGPTSTM IGSRITLMD ENTYOMFOLI GNEFTFDYD SV NIIGCNGLA LYFVSNADG GKAQYKGPNKA GAKYGTGTYC ACOPRDVKFI NQGANSIDWQ PSDSDVNGI GNLGTCPPM DIWEANSLST AVTHPCTKL TQHSTGDC GGTTSNDRYG GTCDDGDFD NSTROGNUTF YCPGSGFEND TTKKVTWVTO FHKGSGNRIS ELTRLYVONG KVIANSESKI AGVPSNLSLT DRCKQKVKY NDPPDFTKKG AWNSMSDALE APMVLYMSLW HDHHSNMLW DSTYPTDSTK LGSGQGSCST SSGVPADLEK NVPNSKVAFS NIKFGPIGST YKSDGNTPTN PTNPEPNTN ANPNPGTVDQ WGQCGGSNYS GPTACKSGFT CKLNIDFYSQ CQ
SEQ ID NO : 39	MYSAAVLATE SELLGAGAQO VGTLLKTESHP PLTIQKCAAG GTCTDDEADSV VLDANWRWHTH STSGSNMNCYT GNTWDTSCLP DAATCTANCA FDGADYEGTY GITISGDSLR LSFVFTGNSV SRTYLMDSSET TYKEFALLGN EFTEDVDSK LPCGLINAGL FVPMADGGM SKYPRINKAGA KYGTGTYCDAQ CPDMKFGVSG GANNEGWP SNSANSGTGN IGSCCSEDEV WEANMSQAL TPHTCTVDQ TACTGDDCAG NTGVDADGC

TABLE 7-continued

SEQ ID NO.	Amino acid sequence
SEQ ID NO : 40	DFNPYRGMNT TPGSGKTKID TIKPFSVVTQ FITDGGTEG TLEIKRPFY QDDVYEQPN SDISGVSGNS ITDDFCTAQK TAPGTDYFS CKGGNAAMGK KMADGMVIALV DSSSPVVAS SPRAVVAQW YGNCITGKTC SEGTCCVYQNP WYSQCVCASS
SEQ ID NO : 41	MFRAAALLAP TCLAMVSGOQ AGTNTAENHP QLOSOOCTS GGCKPLSTKV VLDSNMWRYH STSGTYNCYT GNEWDTSLCP DGKTCANCA LDGDADSYG LKFKTGSMN LMWSDPYPT KDASTPGVSR GSCATTSGVP ATVEAGSGA VTFESIYKG PIGETFKAPA KYGTCYCDSQ CPDKDIFGTL EAVNGWHTT GSNTGNGSYG TCCEMDTNE ANNDAAATP HPCITTSQTR CSGDIDCARNT GLCIGDGCDP NSFRGDKTF LGKMTVDTIS KFTVTVQFL TNDNTSTGL SEIRBLYQN GKVLYQNSVAN IPGVDPVNSI TDNFCAQKQT AFGDINWEAQ KGKHKOMGEA LGNGMVLALIS IWDHAANML WLDSDPDK DPSAEGVARG TCATTSVGPVS DVESQVENSQ VVFSENFKFGD IGSTFSGTSS PNPPGGSSTS SPVTTSPPTV PGTPVQWQ QCGGIGSYGS TTCAСПTCH VLNPIYSSQY
SEQ ID NO : 42	MYRKLAIVSA FLATARASQA CILQSETHPP LTWKQCSSGG TCTQOTGSSVY IDANWRWTHA TNSSTINCYDG NTWSSTLCPD NETCAKNCLL DGAAYASTYG VTIIGNSLISI GFTVTOSSQKVN VGRALYIMAS DTIQEFTLLI GNFRSFDYD SQLPCGNGA LYFVSNADAG GVSKYPTNTA. YPANKAGAKY GTGICDSCQP RDLKPFINGMA NVEGWEPAN DANAGTGHNG SCAAEMLDWE ANSISTAYT HPDCDPGQVN CTGBSCGGTY SSDRGTTGCD PDGCDENSYR QENKTFQGPV YGPGSSPEFL TTKKLTVTQ FETSGAIIYR YVQNGTTEQO PNAELGSGYV NLNLDNYCTA EEAFFGSSSF SDKGGBTQPK KATSGGMVILV MSLUWDDTYAN MLWLDSTPT NETSTPQAV RGSCTSSVSEV PAQVEOSOPN AKVTSNKF GPIGSTGNPS GGNPPFGNRG TTTRPRANT TGSSPGQPS HYGQCGGIGY SGPTVCAST TCQVLPNYYPS QCL
SEQ ID NO : 43	MPSSTDYKK LILLASFLSA SQAQOVGTSK AEVHPSLTWQ TCTSGGSCTT VNGKVVDAN WRWVHNVGDY NNCYGTNTWD TTLCPPDDET ASNCALEGAD YSGNHYGVTIS CNSLRINFTV QASQKNGSR LYLMEDDSTY KMPFLNLQHF TTDVDESNLP CGLNQGAVFYV SMDDAGGMAR YPANKAGAKY GTGICDSCQP RDLKPFINGMA NVEGWEPAN DANAGTGHNG SCAAEMLDWE ANSISTAYT HPDCDPGQVN CTGBSCGGTY SSDRGTTGCD PDGCDENSYR QENKTFQGPV YGPGSSPEFL TTKKLTVTQ FETSGAIIYR YVQNGTTEQO PNAELGSGYV NLNLDNYCTA EEAFFGSSSF SDADGCDENS YRQNTSFYQ AGLTNTNKK FTUVVQFTIN DGTASGTLKE TRPYQVNG VIPNSQSTIA GVPGENSITDS FCAQKTAGR DTNEPATKGG LATMSKALAK GMVLYMILW DHTANMLWD APYPATKSPS APGYTRGSSC ATSGNPVYDV ANSPASSVTF SNIGKGPINS TYTGSGAAPS VPGBTIVSSA PASTATGAG GVAKTAQGG SGYSGATACV SGSTCVALNP YISQCC
SEQ ID NO : 44	MFPAAITLPAF SLFRAVYQGO VGTQLAETHP RLWQKCTRS GGCQTOQSNGA IVLDANWRWV HVNGGYTCY TGNTNTSLC PDGATCAKNC ALDGANYOST YGUTTSNEDL KTKFVTOSEI ESDTQYOLNAG PLNQEFETDV DVSQLPQLGN GAVYFZAMDA DGGNKEPUN AGAKYGTGYC DAQCPDQIKF INGVANEVGW TPSSNDVNG AGOGSCSC MDWEANKLS AAYTPHPCSV DTONTRGTD CGIGARYSSL CDADGCDENS YRQNTSFYQ AGLTNTNKK FTUVVQFTIN DGTASGTLKE TRPYQVNG VIPNSQSTIA GVPGENSITDS FCAQKTAGR DTNEPATKGG LATMSKALAK GMVLYMILW DHTANMLWD APYPATKSPS APGYTRGSSC ATSGNPVYDV ANSPASSVTF SNIGKGPINS TYTGSGAAPS VPGBTIVSSA PASTATGAG GVAKTAQGG SGYSGATACV SGSTCVALNP YISQCC
SEQ ID NO : 45	MKGSISYQI KGAIILSAIL NSVAQOYGT LTAETHPALT WSKTAGKCS QVSSESVIDA MPXVHSTSG STNCYGTNW DATLCPDDVT CAANCAVADGA RROHLRVUTS CNSLRINFTV TASONKNGSR LYLLENDTYT QKENLNLQHF TTDVDESNLP CGLNQGAVFYV DMDADGGMAR YPTNKAGAKY GTGICDSCQP RDLKPFINGMA NVDGWTPSN DVNSLGNHG SCAAEMLDWE ANSISTAYT HPDCDPGQVN CTGBSCGGTY SSTDRTGTTGCD PDGCDENSYR MGVTNFEGPQ ETIDTKSPFT VWTQPLTNDG TSSTGTLSEK RPYVOGKVI GNPOSTIVGV SGNSTSDMRC NAQKSAFGDT NEPSKHGMAI GRGAGLADM VLVMSLMDH ASDMWLSDH YPTNATSTP GAKRGTCIDIS RPNTVTESTY PNAVVIYNSI KTKGRNINSTFT GGTTSSSSSTT TTTSKSTSTS SSSKTTTTV TTTTSSSSSTG TGARDWQCG NGWWTGPTIS VSPYTNKFG DIGSTVSA
SEQ ID NO : 46	MERTAALTAP TLAAVVLQGO VGTLLTAENHP ALSIQQCTAS GCTTQOKSYV LDSNWRWTHS LPVHTHNTCTG NAWDASLCPD PTTCATNCAL DGADYTSYTG ITTSQNAITL RFTVNGPYSK NIGSRVLL DADHVKMFL KNGQFTEDV DMSGLPQGNG ALYFSEMPAD GKAHAHTSNK AGAKYGTGYC DAQCPDQIKF INGEANILDW SASATDANGA NGRIGACAE MDWMEANSEA TAYTPHVCRD EGLYRCGSGTE CGDODNRYGG VCDKDQGDFDN SYRNGDKNFL GRKTIDTWW KITVTVQFTT DDNTSSGNNLV EIRVYVQG VTQYQNSSTP PSLSQTNNSIS DDFVCFNAQK TAPNDTFSFQ QKGGMANMSE

TABLE 7-continued

SEQ ID NO.	Amino acid sequence
SEQ ID NO : 47	FGDNQYNTN GGFHKMGDAM ANGMVLIMSL WSDHAHMLW LDDSYPLDKS PSEPGVSRGA CATTGDDDD VVANHPNASV TFSNIKYGPI GSTYCGSTPP VSSCNTSAPP VTTSTSSCPT TPTGPITGTVP KWGCGGNGY SGPTCVAGS TCYTSDWYS QCL
SEQ ID NO : 48	MYORALLFSA LLSVSRAQOA CTAAQEYVPHS LTWORCEASG SCTEVAGSVV LDSENWRWHTHS VDGTYNCYTG NEWDATLCPD NECAONCAV DGADVEATYG ITENGDSLTL RFTVGSNVS RVYLMEDDET YQMFDLANE FTFDGVSNL PGGLNGALYF TSMDADGLS KYENTAGAK YGTGTCDSOC PRDIKFINGL GNVEGWESD SDANACYGM GTCCPBMDIW EANSISTAYT PHPCDSYBOT MCEGHSGGT YSDRIGGTC DPDGDFNSY RMGNTSFVGP GALIDTSFQD GSLSERFLADG VQGESEVNTA PTEGQZADP DVPSQSYAR GTVESEHADA SVEVSNTAEG PIGSTEF
SEQ ID NO : 49	SIWDDHHSSM MWLDSYYSPD CSLLTETKPA LTWSKCTSG CSVYQGSTI DANWRTHQV SGSTMCHTGN KWDTSVCTSG KVCAEKCCVD GADYASTGYI TSSNOLSLIS FVTKGVTGNG ENYOMPLQI GNEFATPQD SNIGCGLING LYFVSDADG GKAYVPGNKA GAKYGTGCD AQCPRDVIEI NGQANSQDWG PSKSDDYNGI GNLTGNCPEM DIMEANSTI ATYPHPCTKL TOAHCTGDSC GFTFSNDRYG GTCIDADGCFD NAFQGKNTF YGGSGFND TTKKVTVTQ FHKGNSNRLS EITRLYVONG KVIANBSKI AGNPQSLTS DCFTQKVKV GDIDDFAKKG AWNGMSDALE APNVLYMSWL HDHHSHNMLW DSTYPTDSTA LGSGVGCST SSQGPADLEK NVPNSKVAFS NIKTGPGST YNKEGTQQPQ TNPTNPNTN PTNPGBTQW GQCGGTGNYSG PTACKSPFTC KKINDFYSQC Q
SEQ ID NO : 50	MERTAALTAF TLAIVLIGQO VGTLLAENHP ALSIQOCTAS GCTTQOKSTVY LDSENWRWHTHS TAGATNCYTG NAWDSSLCPN PTTCATNCAI DGADYSGTYG ITISGNSLTL RFTVNGYSE NIGSRVYLL DADHYKLENL KNOQFPTFDV MDWEANSAAD GKKAAHTGNN AGAKYGTGYC DAQCPHDITKW INGEANTLW SGSATDPNAG NGRYACAGM DTMWEANSAAD GKKAAHTGNN CGDBDNRYGG VCDKDGCDEN SYRNGDKNFL GRGKTDITW KITVYQTOIT DDNTPTGMLV EIRRVYQDG VTYQNSFSTP PSLSQXNSIS DDECYAQKTL FGDNQYNTN GGTEKMGDLSI ANGMVLIMSL WSDHAHMLW LDDSYPLDKS PSEPGVSRGA CATTGDDDD VVANHPNASV TFSNIKYGPI GSTYGGSTPP VSSGNTSVPP VTSITSSGPT TP1GPTGTVP TGQCGGIGY SGPTCVAGS TCYTSDWYS QCL
SEQ ID NO : 51	MYOKLALISA PLATAQASA CTLOAETHP LTWQKSSG TTCTQGTSTV IDANWRWHTHS TNSNSTCVD NTWSSTLCPD NETAKRNCLU DGAAATASYG VTTSADSLSI GFTVOSAQN VGRALYIMAS DTTOEFTLL GNEFSDYDV SOLPGLNGA LYFVSDADG GFTVSKPTNTA GAKYGTGCD SQCPRDILKEI NGQANVNGA PSSNNTANTCI GHGSSCCSEM DIMEANSTI SE ALTPHPCTTV GOEICEGDSC GFTFSNDRYG GTCDCPDGCDW NPYRLGNTSFID TTKKLTVTQ FETSGAINRY YVQNGVTHQO PNAELGHYSG NSLDIYCAA EEAFFGSSSF SDKGSLTQFK KATSGMVLW MSILWDYYAN MLWLDSTYPI DTSSTPGAV RGSSTSSGCV PAQLENSPN AKVYNSNIKF GPISGTGNPS CGNPGCGNPP GTTPRPTA TGSPPGPTQ HYGOCGQGIGY IGPTVAGS TCYLNPTVY QCL
SEQ ID NO : 52	MTWOSCTAKG SCINNKNGKLV IDANWRWHTHS KEGYDNCYTG NEWDATACPD NKAACAANCV DGDYSGTYG ITAGGNSLKL KFTIKGYSST MIGSRTYLMK DDTYTMFKF TGNQEPFFD DVSNLPGFGN GALIVFMSA DGLLKKSTN KAGAKTGTY CDAQCPDRLK FINGEGNVEG WKPSNDANA GVGCHGSSCA ENDIWEANSTI STAVTHSCS TIEOERGCDG GCGTYSAADR YAGVCDPDGC DFNSTRMGVK DFYRKGTVD TSKKEFTVTO FIGTGDANEI KRFYVQNGKT IAQPAAVPG VEGNSITKF CDQOKAVFED TYTFDKGGM ANMAZALANG MVLVMSLWD HYSNMLWDS TYPIDKNPTD DLGTRGRCCE TSSGVPADVE SOHADATVY SNLKFGPLNS TRG
SEQ ID NO : 53	MASATSOVY RSALISAPL PSITOAOQIG TTETTHPSM TWETCTSGS CATNOGSYVM DANWRWHTHS CATTGNCYTGN TWDDTSICD TD ETCACTECADV GADYESTGVV ITSGQRLIN FVTQNSNGAN VGRSLYIMAD NTHYQMKFL NQEFPTDVIDV SNLPGLNGA LYFVSDADG GVSKYPPNKA GAOTGVGCD SCOPRDLKFI QGOANVEGT PSSNNTGTL GNTGSCCAEL DIWENSTI SO ALTPHPCTDA TNTCTGDAC GTGTYSSDRYA GTCDPBGCDP NYRGKTFIDT YGPGBTQI LDGTDITGEL STTGTQVQN GTVYQDPSD ISGTIGNAIN ADYCAENTV FDGPGBTPKH GFPSAMFEM STGMVILVMSL WDDYYADMILW LDSTYPTNMS SSTPGAVRGS CSTDSEYPAT IESTSPDSTV TFSNIKVGPI GSTFSSSGS GSSGSGSSGS ASTSTTSTKTI TAATSTTAV AQHYSGCGQ DWGPTTCVY PYTCQVQNAV YSQL
SEQ ID NO : 54	MKAYPEYLVA ALPLJGLATA QVQGKQFTE HPKLSWKKCT GRANENTVNA EYVLSNPRW LHDSGGRNY DGNKNTSACS SATDCASKCQ LDGANYGTTT GASPTSGDALT LKEVTKHEY TNIGSPRY LM NGASXYQMF LMNNEFADY DISTVREGLN AALYVYAMEE DGMASYSNN KAGAKYGTG YVGGKANL EG WPTSTNDANA GVGPGVCCA EIDWWEIAH SPAFTPHACK TNKHYCERD NGGTYSEDR FAIGHCDANGC DYNPYRMNT DFYKGKHTVD TSKKEFTVSR FEENKLTOFF VQNGQKIEP GPKWDGIPSD NANLTFEECS AQRQAFGDRD RFAEYGGFAQ LNSALRMPMV LVMSIWDDH ANMLWLDVVY PPEKEGQGGA ARGDCPQSSG VBAEVESQYA NSKVVYSNR FGPGSTVAV
SEQ ID NO : 55	MFSKRALTGS LLAGAVNAGC YOTQQTETHP QMTWQSTSP SSCNTNGEV VIDSNNRWHV DKDGYNCYT GNTWMTLCP DDKTCZAANCV LDGADYSSTY GITTSIGNALS LQFVTOSSGK NIGSRVYME SSTKXHFLDL IGNERAFDVD LSKLPGCLNG ALYFVMDAD GMKAYSTNT

TABLE 7-continued

SEQ ID NO.	Amino acid sequence
SEQ ID NO : 55	AGAEGTGTGIC DSQCPRLIKP INQGQNEVGW TPSTNDANAG VGGLGSCCSE MDVWEANSMD MAYTPHPCT AQQHSCNADE CGGTYSSSRY AGDCDUDGCD WNPFRMKGKD DH2ANMLWID GMVLVMSLWD STYPUVADAS SPKQRGTCAT TSGVPADVE SDAVASITY SNKPGPTGA TY
SEQ ID NO : 56	MPPAAALLSF TLLAVASAOQ IGTNTAETH SLTVSQSTS GGCTSSTQI VLDANWRWLH STSGYNTCYT GNOMNSDLCP DPDTCATNCA LDGASYESTY GISTDGNAVT LNFVTOGSOT NVGRSYRILS DDTHYOTPSL LNKEFSEFDV ASNIGCGING AVYFTOMDAD GGHLSKYSSNK DSQCPDLIKP GMTCGDTENSC RMGNETFLKP CMTDITMKK TIVTOGTDD NTANGTSDV RLYVQGTYT QLNSVANOPN IPKVNSITDS FCTQKTERG DKGDGTDENSC DQYDFTGIGG LSQNGKANSD MVLVMSLWD YDAEMLWLD SNTSHSAAAAT TATSSAAAAT GSPVPAVGQC GGQSYTGSTT CASPVCKVS NAYSQCLPA.
SEQ ID NO : 57	MKRALCASL LLAAVAQVQ GTNEPEVHPK MTWKKGSSGG SCSTVNGEVV IDGNMRWTHN IGGYENCYSG NKWTWYCSTN ADCATKCAAE GAKYQRTGYV STSEDALTLPK FVQCNSSSKRN VGSRMLMNG ANKQMFPLK NNEFAFDVYL SSVEGMNIA LYFVPMKEDG GMSTEPNNKA GAKYGTGTYC AQCARDLIEPI GKGKNGIWMQ PSSTDSSAGI GAAGCACAEI DINESNKAF AFTPHPCENN EYHVCETPNC GGTADDRYG GGCDAANGDGY NPTRMGNPDE YGPGBTIDTN RKFVTSRFE KRDGLIEBGTN ELNEQFCTDQ FTVDPERNRF NEVGGWMSKLN AAYEIPMLV MSIWSDHFAN MLWLDSTYPP EKAGQPSAR GPCPADGEPP NGVYMQPNM KVIMSNVRFG PIGSTYQDV MSLWDDHYSN MLWLDSTYPP DKEGSPAAK GDCPQDSGPV SEVEQSPGA TVVWSNTRFG PGYSTNV
SEQ ID NO : 58	MQLTKAGVFL GALMGAAAQ QVGTQTAENH PKMTWIKCTG KASCTVNGE VVIDANWRWL HDASSKNCYD GNRWTDSCRT ASDCAAKCSL EGADATKTYC ASTGKTYGK ASTGARDLKP DAQCARDLKP VGGKANIEGW KPSNDANAG VGPYGACAAE IDWESNAHA FAFTPHCTD NKYHVQDSN CGGTYSDDRF AGKCDANGD GDIPTRLQGTD FVKGKTVTDT SKFKFTVTRF ERDALTOFFY QINRKINDPS PALEGIPATG AITAETYCTN ENVFRDRNRF DEYGGWSOLQ QALSILPNVLF YKSDGTTPNP TPNPPAGS VDQGOGCKN YSGPTCKSP FTCKINDPSY SOCQ CSINTETKDA LTWSKCTSG CSDYKGSVVI DAWBWTQHOT SCSINTCYGN KWDTSICLTDG KTCAEKCCLD GADYGTGTYI TSSGNOLSLG FVTPNGPSKIN IGSRTYLMEN ENTYOMFQLI GNEFTFDYDV SFGCGGNGA PHFVNMDEDG GKAYKSGNKA GAKYGTGTYC AQCPRDVKPI NGYANSECKW PSDSYNAGY GNLTGTCPEM DIWEMANSLST AFTPHPCTKL TOHSCTGDS C GFTYSDRYG GTCDCADGCFD NATROGNTIF TTKKMVTUO FHKGENGKELI EITRLYVONG KYIANSKSKI AGNPASLTS DFCRKQKSVP GDIDDESKKG PWNGMSDALS APMVLVMSL HDHHSNMLWL DSTPTDSTK VGSQSGCAT TSGKPSDLER DVPSKVSFS NIKGPIGST YKSDGTTPNP PASSTTGSV TPNPPAGS DowGOGCKN YSGPTCKSP FTCKINDPSY SOCQ
SEQ ID NO : 59	MYORALLFSA LATAVSAQQ GTOKAEVHPA LTWOKCTAG SCTDQKGSVY IDANWRWLHS TINSGAITLED ANWRVHGWN TSTNCGTGN TNEWAELCPD NEAACANCL DGADYGTGTY VTAGDSSLKL NEVTSANLPS RLYLMEDDT YOMNLINNE FTFDVSILN PCLGNGALYF VSMADAGGLS KYPGKAGAK YGTGTCDSOC PRDLKIFINGE ANVEGWPSPD NDKNASVGGY GSCCPMDIN EANSISIATY PHPCDGMET RCDGNDGCGT YSSTRYAGTC DPDGDDENSF RMGNESFPGP GLVDTKSPV ALIVSINTDH ASSMWMLDST YPVDADSTP GVARGTCHG LGDPETVESQ HPDASTYFSN IKFPIGSTY KSY EDIFSKHGGI EGMCALKNM ATAGAQOIGT YTAETHPSL WSTCKSGSC TINSGAITLED ANWRVHGWN TSTNCGTGN WNTAICLTDIA SCQAQDALDG ADYGTGTYT TGSNLSLNP VTGSNGSRT YLMADNTHYQ FDLINQEFT FTVDFVSILPC GLNGLALYFV MDAGGVSKY SNRYAGTCDP DGCDNPFL DGCDNPFL VEGWTSTNN SNTGIGNHG CCSELDWEA NSISEALTPH PCDTPLHTYC TADCGGTYT AELSARGETA SFTNHGGKRN MGSALEAGMV LVMSLWDDYS VNMLWLDSTY PANETGTPGA ARGSCPTSG SNTVYESQG VGPENSTPSG GTSTGGSNTT TAGTGSTKA STTSTSSTST GTGVAAHWQ CGQQGTGPT TCASGTTCTV VNPYSQCL
SEQ ID NO : 60	MRTAKEATLA ALVASAQQQ ASLTERHP SLISWNKCTAG GOCQTVQASI TLDSNWRWLH OVSGSTNCYT GNKWDTSIC TAKSAQNCC VDGADYTSY GIFTNGDSLQ LKEVTKSGHS TNVGSPTYL DMENKYQOTPE LIGNEFPTDY DVSNIIGGLN GALVYPSMDA DGGSLRYPEN KAGAKYGTGTY CDACOPRDJK FINGEANLEW WTGSTNDPNA GAGRYGTCSS EMDIWEANN AVAFTPHCT IIGQSRCEGD SGGTYSNER YAGVCDPDGC DENSYRQENK TEYKGKMTVD TTKKLTIVTO FLKDANGD LG EIRRFYQDG KLIPINSTSI PGVEINSITQ DWCRDQKVAF GDIDDENRKG TVAGLPGAGN GENGGNTPP PTITTSAPA TTTTASAGPK AGRNQOQGGI GTFGPTQOEE PYICKLNDW YSOCL
SEQ ID NO : 61	

TABLE 7-continued

SEQ ID NO.	Amino acid sequence
SEQ ID NO : 62	MMYKKPAALA ALVAGASAOQ ACSLTAENHP SLTWKRCTSG GSCSTVNGAV TIDANWRATH TVSGSTNCYT GNWDTSLCT DGKSCAQITCC VDGAIYSSTI GITTSGKVNN LKFTVKHQV TNGVTKLNL ENDKTYVLL DNGEFTDVI DVSNGGLN GALIFMSMDA DGGNSKYSQN KAGAKYGTGY CDAQCPDLK FINGEAVNGN WTPSTNDANA GFGYGSCCS EMDVWEAMIN DTAATPHBCT TVGOSCEAD TCGSTYSSDR DEMYARQDK TFVKGKMTVD TNKMKUTVTVQ FHNSAGVLIS EIRFVYQDG KLIANAEISKI PGNPNSITQ EYCAQKVAF SNTIDDEFNRKG GMQMSKALA GMVLYMWSW DDHYANMLWL DSTYPIDAOQ APEAERGACP TISGVPAEIE AVOPHSNVIF SNTBEGPPISS TVPGLGSNP GNPTTIVVPP ASTSTRPT STSSPVSTP QPGGCTQK WQGQGGIYT GCINCYAGTT CTQINWYSQ CL
SEQ ID NO : 63	MASLSLISKIC RNALILSSYL STAQQGQVGT YOTETHAPSMT WOTCNGGSC STNGGSTVLD AMWRWHTQ SSSMCYTGK WDTSYCSTD TSQEVPLNF VTSNSNGKVN GSRYVMMADD THYEVYKLN QEFTEVDVQ KLPCLGNGAL YEVVMDADG KXAYPGNKA VSKYPNKAG QCPDLKFQI QGANYEVGMW QPSKSDTNAGI GNMGTCQPEN DIVEANST AVTPHPCTKL IWEINSLQA LTPHPCDTQ NTLCGDAAG GTYSSDRTSG TCOPDGDFN PYRUGNTTFY QPKTLDWK PITYVQDFIT DDTGNSGTLT EIKRFVYQDG VTPQPSADY SGHSGNTINS EYCTAENTLF EGSGSPAKHG GLAGNGEAMS TGWLVLVMSW DYYANNMLW DSNYPNTMEST SKPGVARGTC STSGSVPESE EASNPNSAVYA YSNIVKVGPIG STFKS
SEQ ID NO : 64	MYRALATASA LIAAARVQQY CSLTPETPKA LSWSKCTS SGSVQSVTI DAWNRWTHOL SGSSTCYTGK WDTSYCSTD KVCAEKCCID GAEYASTYGI TSSGENOLSIIS FVTKGAYTGNL ISGRSTYLMED ENTYQMRQFL GNEFTEDVYD SNIGGCLNGA LYEVNSDADG KXAYPGNKA GAKYGTGYCD ACOPRDVTFI NGANSDDWQ PSKSQDNAGI GNMGTCQPEN DIVEANST AVTPHPCTKL TQHSTCGDCS GGTYSNDRTG GTCDADGCDP NAYRQGNITF YGPMSGFNUV TTKKVTVTQ FHKGNSNGLIS ETIRLYVQNG KYLIANESSEKI AGVPSSLTP ERCAQKVKV GDTDDFAKKG AWSEMDSALW DSHHSNMLW DSTYPTDSTK LGHQGQSCST SSGVPADLEK NVPNSKVAFS NIKEPIGST YKEGYPEPTN PTNPNTNTPN TNPGBTVDOWA QCGGTYNSGP TACKSPETCK KINDPSQCQ
SEQ ID NO : 65	MFPKSSLLVL SELATAYAQO VGTQTAEVHP SLNWARCTSS GCTNVAGSVT LDANWRWLHT TSGYTNCTG NSWNTLCPD GATAQNCAL DGANTQSTCG ITISGNALTL KEVTQGEONL IGSRPTVLMAS ESRVEMGILL DNEFTEDVYD SNLPCGLNGA LYEVNSDADG GMANQPGNKA SQPDRDIFKI INCEANAVCMN GSPNPUTNACT CNGACGNCM DIVEANST AVTPHPCTKL QLSPCGLNGA CTRNRYGTC GAKYGTGYCD SCOPRDIFKI INCEANAVCMN GSPNPUTNACT CNGACGNCM DIVEANST AVTPHPCTKL QLSPCGLNGA CTRNRYGTC DPDGCDENSY RMGDKTYGP GETGVDRTRSK FTVVTOPLTN MNSSTGTLSE IRFLYVQNGR VYQNSKVNIP GMSNTLSDIT TGFCDSDKTA FGDTFSRQNQ GMNSAMGQAL GAGMVVLWS WDDHAANMLW LDNSYVPDAD PSKPGIARGT CSTTSQGKTD VEQSANSV TFSNIKRGDI GTTYTGGSVT TTPGPNPGTTT STAPGAOTK WGQCGEQGWT GPTRCESGST CTVNONTSQ C1
SEQ ID NO : 66	MFRKAALLAF SFLAIAHGGQ VGTMQAENHP SLPSOHCTAS GCTTSSTSVY LDANWRWLHT TGTGTYNCYT QWDASICPD GVTCAKACAL DGAQYSGTYG ITISGNALTL QFVKGTGIVS RYVLLQDASN YQLFKLNLNQ FTFDVMDSL POGLINGAVYL SQMDQDGVS REPTNTAGAK YGTGTYCDSOC PRDIFKINGE ANVAGNTGSS SDPNSCTEDVN GTCSCEMDW EANSVAATAV PHPCSTYQOT RCTGADCGOD ANRKGVCDP DGDCDENSFRM GDQFLGKCL TVTDSRKFV YVQFLISDDGT SGNLAETIR PYQDGKVPN NSTOEWCDA QKVAFDLIDD FNRKGENAOM SKALEGPWVL NRFADQGGLK OMGAULKSMQ YAAQMLWHD AANMLWHD YPTTADASKP GVARGTCPNT SCKVYDTEQO SGSVATYTSN IKWDLNSTR SGTASNPQGP SSSSPGPSSS SSTAGSQT QPSSGQVW QCGGIGYSG ATGCVSPYTC HVNMPYTSQ Y
SEQ ID NO : 67	TETHPRLTWK RCTSGGCST VNGAVTIDAN WRWHTIVSGS TNCYTGNEWD TSICSDGKSC AQTCCTDGDAD YSSTYGLITS GDSLNLKFVT KHOHCTNNGVS RYVLMENDTK YOMFELLIGE FTEDDVDSNL GCGGNGALYV VSNDADGMS KYSGNKAGAK YGTGTYCDAOC PRDIFKINGE ANIENWTPST NDANAGFREY GS CCSEMIDW EANNMELAFT PHPTIIGQS RCEGNSCGT YSSERZAVC DPDGCDENAY RQDKTFYKG VMTYDFTKKM TVTQFHKNIS AGYLVSEKRFV YVQDGKILIAN AESKIPGNPG NSTOEWCDA QKVAFDLIDD FNRKGENAOM SKALEGPWVL VMSVSDTTOIS NMWLDSTYD TPTSQPGGCTV SNTVFSNRPV GFTGTVPLG DGSTPSNPTA TVAPPSTTT SVRSSTTOIS TPTSQPGGCTV YTKWGOGCCII GYTGCNTVVA GTTCETLNPW YSQLC
SEQ ID NO : 68	MFHKAIVLAP SLYTIVHGOO AGTOTAENHP QLSSOKCTAG GSCTSASTSV VLDNSNMWVH TTSGYTNCTG GNTWDASIC DPVSCAONCA LDGAIYAGTY GIFTSGDAIT LKFVFTGSNG SRVYLMED NYQNEKLMMQ EFTEVDVSN LPGCLNIAVY FVQMDODGTT SKFDPNINKAGA KEGTGYCDSO CPQDIFKINGE EANIVWDTAS AGDANEHTGS EGTCQEMDI WEANSTISAY TPHPCTYEQ TRCS3DCCQ GSDFENGICD PDGCDENSFR MGNTBFYGRG LTVDFTSQFPT IVTQFISDGG TAGNLAETIR RFVQNGKVI PNSVQQTGI DVNSTEDF CTOQFTVFGD TNNFAAKGGI KONGEAVING MYLALSLWHD YAQMLWHD DYTPTADPSQ PGTARGTCTP TSGVPSQVEG QEGSSSVIYS NIKRGLDNST FTGTLTNPSS PAGPPVTTSP SEPSOSTOPS QPAOPTQPG TAAQWAQGG NGFTGBPTVCA SPFTCHVLPN YYSQCY
SEQ ID NO : 69	MFRAAALLAF TCAIAWMSGQ AGTNTAENHP QLSSQOOTS GGCRPLSTKV VLDNSNMWVH STSGYTNCTG GNEMWTSCLP DGKTCAAANCA LDGADYSGTY GIISTGTALT LKFVFTGSNG SRVYLMADDT HYQLLKLINQ EFTEVDVSN LPGCLNIALY LSAMDADGGM SKYFGNKGAGA

TABLE 7-continued

SEQ ID NO.	Amino acid sequence
	KYGTGTYCDSQ CPIDIKFING EANVGNWTE GSNTGIGSYG TCCSEMDIVE ANDDAAFTP HP HPCOTTGOTR CSGDDCART GLCHGDGCD FNSPFRNGDKT FLRGKMTVDT SKPFTDVDF LTNDNTSTGP LSEERRIYIO NGKVITNEYVA NIPGVDPVNS ITDNFCAOK TAFCDTNWPA QKGGGLKQMGTE ALGRNGMAMT SLWDHSDYPT KDPASGPVAR GTCAVTPSVPN SDEVESQPNs QVVFNIKFG DIGTSFGTS SPNPFGGSTT SSPVTTSPTP PPTGPTVPOW GCGGGGSG STTCASPYC HVLNPNYSQC Y
SEQ ID NO : 70	MMMKOYLOYL AAALPLVGLA AGORAGNETP ENHPPLTWOR CTAPGNCOTV NAEVVIDANW RWLHDNDMON CYDHNONTNA CSTADCAEK CWSIEGAGDARYL GTVGASTED ALTFKFTVPR YLMAGDFTYKQ MEULNGNELA FAYDIDWNEA FVNSALYFVE MEEIGGMASY PSNOQAGARYL TGKCDACQAR DLKEVGGKAN LEGWKSTSD PNAGPGYPS CCAGIDWNEA FAYDIDWNEA FVNSALYFVE MEEIGGMASY EDRFAGRKDA NGCDYNPFRM GRPDIFYGKX TLDTSKRPTV VSREFBNKLX QYFTQDGKX ELPPTWBMG PNSSEITPEL CSTMEDVEND RNRFEEVGGF EQJNNALRVP MVLVMSIWDD HYANMLWDS IYPPEKEQGP GAARGDQPTD SGVPAEVEAQ FPPAQVWNSN IRFGPGSTY DF
SEQ ID NO : 71	MYRSATELTFP ASLVLGQOYG TYAERHPSM PIQVCTAPGQ CTRESTEVYL DANWRWHTH NGYTMCYTG EWAAATACPDG ATCAKNCAYD GADYSCTYGI TIPSGAIDL QPVFKNDQ NVGSRYLMA SSDKTKFLNL LNKETFDVQD VSKLPGCLNG AVYFSEMLD GGKCSFSGNK AGAKYGTGYC DSQCPDQIKF INGEANEGW GGADENSGTG KYGICCAEMD IMEANS DATA YTPHVCYNE QTRCEVDCG AGSBDYNSIC DKDGCDFNSY RLGNREFFYCP GRTVDTDRP TIVTOVTTD GTDSENKLX HRYVQDGNV IPNSVTEAVG VDQTMFISEG FCBQOKSARG DNNYFGQLGG MRANGELESKK MYLVLSIWDD HAVNMMLWD STYPDADPEQ PGYARGRCDP ADGVPAIE AHPDVIYI NIKGAINST FTAN
SEQ ID NO : 72	MYRTLARASL SLYGAARAOQ VGTSTAENHP KLTWQTCTGT GGTNCSNKG SVYLDSTWHR AWHNVGTYNC YTGNISNSTQY CPDGDSCTKN CAIDGADYSG TYGFTTSNA LSUKFTVQD PSSNIGTQ FMEDTKYQ FMNLSKTKLNGE LNKETFDVQD VSKLPGCLNG AVYFSEMLD GGKCSFSGNK NKAGAKYGTG YCDSQCPDII KPTNGKANE GWNPSADPN GGAKGKACG PENDIWEAS DITYAHTHPC RGVHQECSD AAEGDGSNR YDQCDKDGC DFNSYRMGKX DEYGPATL TTKKMTVITO FGLESSSLLSE IKRFYVQNGK VYKNQSOAVA GVTGNITES FCTAQKARG DTSSPAALGG LNENGASLAR GHVLIMSLW DHAVNMMLWD STYPDADPS KPGAAARTCP TTSGKPDVE KNSPDATVVF SNIKFGPIGS TFAQPA
SEQ ID NO : 73	MYOKLALISA PLATAROSA CTLOAETHPD LTWQKSSG FSSNIGTQ DTTQEFITL GNEFSFVDV SOLPGCLNGA LYFVNDADG GVSKYPTNTA DGAATASTYQ VTTSADSUJ GFTVOSAQN VGANLTMAS DTTQEFITL GNEFSFVDV SOLPGCLNGA LYFVNDADG GVSKYPTNTA GAKYGTGYC SQCPDQIPEI NQGQANVEWA PSSNINANTCI GGHSGCCSEM DIMEANNSLE ALTHPHCTTV GOEICDGDCS GFTISGDRYQ GTCDCPDGCDW NPYRLGNTSF YCPGSSPFILD TTKKLUTVTO FETSGAINRY YVONGVFOO PNAELGHYSG NSLDDYCAA EEAFFGGSSF SDKGSLTQFK KATSGGMVLV MSLWDYTYAN MLLWDLSTP1 NETSSTPGAV RGSCSSTSSCV PAQLENSNSPN AKVVSNIKF GPISGTGNSS CGNPPGQNPQ STGSSPQPTO THYOGCGGIG YSGPPVQASG STCQVLPY YYCQL
SEQ ID NO : 74	MWDSFSIYKT ALLSMSLATS NAOQVGTYA ETHPSLTWQ CSGSGSCTT SGSVVDANW RWHEVGGYT NCYSGNTWDS SICSTDTCIA SECALEGFTY ESTYGVTSQ SSIRLNFTV ASQKNIGSLR VLLADDSTT TPKLNFRFT FDVDSNLPC GLNGALFYV SRF PTNKAGAKYK TGYCDSOCPR DLKPIDGQAN LEGWEPSSD VNAGTGNHGS CCEPMIDVEA NSISATTAH PCDSYQOTMC TGDTCGGTYS DTTDRYSGTC DPQCDENPY RGTNTNYP GKTVDNSKPV TVVTOITHD GTDTGTLTEI RELYVQNGVV IGGPSTYTA ASGNSITESF CKAEKTLFGD TNVTEFHGL SAMGDAQGD MVLSLWDD HAADMWLDS DYTPTSCSS PEYARGCTCPT TTGNAFVYEA NYPSVYVTS NIKFGTLNIST YSGTSSSGSS SSSTLTTKA STTSTSKTTI TTTSKTTT SSSNTAQY QCGGGQGWG PTTCASGTCIKQDYYQL
SEQ ID NO : 75	MYRIKSFIL LSLYNMSLQ KIGKLTPEVH PMTFOKCSE GSCTEIQE VVVDANMWV HSAQGQNCYT GNTWMPCTC DDETCACNY LDGANYESVY GVTISEDSYR LNFVTOSGK NIGSRFLMS NESNYOLFH LGEOFETFDVQD VSNLDCGNG ALYVSMDSD GGSRFPTE PTNKAGAKYV DAQCPDRDIFL TGSANTDGF IPSTMPNT YGNLSSCAE MDIWEANNA TAVTPHPCDT SSQSTSRSDS CGGAASSHNR GICUDPDGCD YNPYRGMNTS FPROPNKMDT NSVITYVTOF ITDGGSSDGK LTSKRLTYO DENVISOSVS TIDTEGNEV NEEROTNOKK VFGDBEDSFTK HGGIAKMGSEA LKDGMLVLVLS LMDDYQANML WLDSSYPTT SPTDPGVARG SCPTTSGPV KVEQNPVAY VVSNIKVGP IDSTYKK
SEQ ID NO : 76	MISRYLAISS LLAARAQOI GINTAEVHPA LTSIVDANW RWLHTTSQY NCYTGNSNDA TLCPDAVTA ANCALDGDY SGTVGTTSG NISLKUNEVTK GANTNVGERT YLMAAGSKTKY YOLLKLUQE ETENYDVSNL PCGUNGLALYF AEMDADGVS REPTNKAQO YGTCYCDAQI PDIKRPINGO ANSVGWTPS NDVNTGTCQY GSCCSSENDW BANK1SAYT PHPSVSDQZ RCTGTDGIG ARYSLLDAD GDINPSYRQG DTGFTGAGLT VDTISKVFTV TOFITNDGT SGTLSERF YVONGKVIPI SQQSKVTCG NSITDSCAA QKTAEGDTNE FATEGGLATM SKALLAKGMVL VMSIWDDISA MNLWLDAPP ASKSPAAVG SRGSASSG VPADVEANSP GASVTSNIK WGPINSTYSA GTGSNTGSGS GSTTIVLSSV PSSTPSTTIG VPKYQOCGS GYTGPCTCIG STCYSMQYY SQQQ

TABLE 7-continued

SEQ ID NO.	Amino acid sequence
SEQ ID NO : 77	M Y R O V A T A L S F A S L V L G O O Y G T L T A E T H P S L P I E V C T A P G S C T K D E T T V Y L D A N W R I T H Y T D G Y T C Y T G N A M N E T A C P D G K T I C A A N C A I D G A E F E K T I G T T P E B G A I R L N F T E P S I V N A M E D K Y R L E N L M K E F I D V D Y S N L P C G I L M A G V Y F S E M D D G M S R F C N K A G A K Y G T G C D S Q C P R D I K F P E G E N S E G E G G E D G N S C T G K Y G T C C A E M D I N E A N L D A T T V P H P C K V T B O T R C E D D E C G G A D A R T E G L C D R D G C D F N S F R L G N K E F T I G G P L P A M G K S L E K M V I V L S I W D H A V Y M N W I L D S N Y P T D A P T K P G V A R G C D F E A G P E T V E A H P D A Y T I Y S N I K G A L N S T P A A A
SEQ ID NO : 78	M S S R Q V Y R A A L I I S I L T A T A N A Q V Q G Y T T E T H P B S L T W O T C T S D G S C T T N D G E Y U V I D A N W R W H Y S T S A T N C Y T G N E W D T S I C T D D V T C A A N C A I C A L D G A T E Y A T G V T T S S E R L A N F V T Q G S S K N G S R L Y L M S D P S N Y E L F K L G Q E T F P D V D S N D F C G L N G A L Y F V A M D A D G T C E S Y E S G N K A G A K Y G T G Y C D S O C P R D L K P I N G E A N C D G W E P S N Y N N T G Y G D H G S C C A E M D V M A N C D V T Q S C G D A C G D S G C G T Y S A S G D R Y S G T C D P D G D Y N P R V L T V D I T D S F P V T Q F I T D D G T S S E T L E K R L Y V Q N G E V I A N G A S T Y S S V A R C E S E K T I L E G D E N V E D K H G G E E G N G E A M A K G M V I V L S I W D M D Y A D M W L D S D Y P T A S T P G V C T V C G Y A R G T C S T D S G V P A T V E A E S P N A V T Y S N I K G P I S S S S S T T K A T S T L K U T K S T T K T T S G S S S T T K A Q A Y Q Q G G Q G W t Q P T C V S G T I C Y T E N A Y S Q C L
SEQ ID NO : 79	M Y R A T A S A L L A T A R A Q Q C T L T A N K P A L T W A K C T S G C S N V R G P T C D V D N W R A H S T S S S T C N V C Y T G N A M N E T A C P D G K T C A D K C C L D G A D Y S G T Y G V T S S G N Q N L M K F V T V G P / F S T N V G S R L X I M E D E N N Y Q M E D L I G N E F T D V D Y L N G A L Y F V A M D G K S P R E S T N I K A G A K Y G T G C D S Q C P R D V K F I P D G D G C D F N P Y Q R G N I T F G W S C M D A L A K P V N Y L M S L H D H Y A M N L W L D S T V P K G S K T P G S A R G S C P E D S G D P D T L E K E V P N S G V S F S N I K R P I G S T N D P D F E N K K G Y T G G S N P D P E E P E E P E P V G T V P Q Y Q C Q G G I N Y S G P T A C V S P Y K N K I N D F Y S Q C Q
SEQ ID NO : 80	E O A G T A A E N H P P L T W O Q C T A P G S C T Q N G A V V U L D A N W R W V H D Y N G Y T N C Y T G N T W D P T C P D D E T C A Q N C A L D G A D Y E G T Y G T S S G S S I L K L N F V T G S N V G S S N L Y N Q I L I D Y T Q S P L Y K I L N Q I S P F D V D N S N L P G C L N G A L Y F V A M D G K S P R E S T N I K A D C B A N V E C G Q P S E M N A L I Q C D H Q S C C A E M D V N W E A N S T N V T H P C D T P G C P M C G D C F N P Y M G N T S F Y G P C H I L D T T K P F I V V T O F L T D D G T D I T S L T S E I K R Y I O N S S N V I Q P O N S I S G V T N S T K T E F C T A Q K O A F G D T D P R S Q H G G L A R M G A M Q Q G M V L V M S L W D D Y A Q M L W L D S D Y P T D A D P T P G I A R G T C P T D S G V P S D V E Q S P N S Y V T Y S N I K G P I N S T F T S
SEQ ID NO : 81	M E P T I A L V S L S E L A I A Y Q Q V G T L T A E T H P K L S V S Q C T A G G S C T T V Q R S Y V L D S N M R W H D V G G S T N C Y T G N T D D I T D S L C P D P T C A A N C A I D G A Y S G T Y G I T S G N A L S L K F V T Q G P S T N G V T N I G S V Y L L S E D D S T Y E M F N L Q N O E F F D V D M S A P C G L N G A L Y F V M G A L F V M D K D G S G R F P T N K A G S K Y G T G Y C D T Q C P D I I K P I N G E A N V L D W G S S N D P N A G T H Y G T C C N E M D I W E A N S M G A A V T H V C T V Q G Q T R C E G T D C G D G D E R Y D G I C D K D G C D F N S W R M G D O T F U G P G K T V D T S S K F T V T V T O F I T A D N T S G D L S E I R R L Y V O N G R V I A N S K T Q I A G D I A Y D S I T D F C N A Q K T T F G D T N T F E Q M G S L A T I G D C A F E T G M V L W M S I W D D H E A R M L W L D S D Y P T D A D S A P G V S R G C P T S G D P T D V E S S P G A T V I S N I K G P I P I G S T F T S
SEQ ID NO : 82	M L S A S K A A A I I L A P A H T A S A W V G D Q Q T E T H P K L A V Q R C T G K G R S S C T N V N G E V I D A M W R L A H R S G Y T N C Y T G E W N Q S A C P N N E A C T K N C A L E G S D Y A G T Y G I T S G N O M N I K E P T K R P Y S T N I G A R T Y L M K D E Q N E M P O L I G N I C E T D V D S O R C G M G A L Y F V S M P Q C G Q G P A G A K Y G T G C D A Q C A R D I K F V R G S A N A G W T K S A S C D P M S V G K Q A C Q M D Y S V C V D D T C G G T Y S E D D I N C G G T Y S E D R Y A G T C D A N G C D F N P F R V G K D F Y G R G K T V D T S K M T V T Q V T V G V S G N O U S E K R F Y O D G K V I A N E P T P G M E W C N T Q K V F Q E E A P P E N E R G G M A S M S E G M S Q G M V L V M S L W M D H Y A M N L W L D S N W P R E A D P A K P G A R D C P T S G K P S E V E A N P N A Q V M F S N I K G P I S T F H A A A
SEQ ID NO : 83	M F R T A L L A F T M A A M V F G Q Q V G T T A R S H A P L A T S Q R C T K G G C S N L A N T K I V I D A N W R W H S T G Y T C Y T G N Q W D A T L C P D G K T C A A N C A I D G A D Y A G T G T Y G T A T G S S S T K L Q F V T G S N V G S R V Y L M A D D T H Y C M P L I N Q E F F D V D M S N L P C G L I N Q A L Y L S A M A D G M A K Y P T N K A G A K Y G T G C D S Q C P R D I K F P E G E N S E G E G G E D G N S C T G K Y G T C C A E M D I N E A N D A A Y T P H P C T M Q T R C S G D C T R D T G L C D A D G C D F M S F R G D O T P I L G K S L T V D T S K P T V V T Q V T I T D G T S A G T L T E I R L Y V Q N G K V I O N S S V K I P G I D P T N S I T D N F P S O Q K T A F G T N Y F A Q H G G L K Q V G E A L R T G M V L A L I S I M D Y A A M L W L D S N T P N K D P S T P G V A R G T C A T T S G V P A Q I B A Q S P N A Y V F F S N I K F G D L N T Y T G T V S S S S V S S S S S S S S T S T S S S H S S S S S S P P T Q P T G V V P Q M Q C G G I G Y G S T T C A S P Y T C H V N P Y Y S Q C Y
SEQ ID NO : 84	M Y Q R A L F S A L M A G V S A Q Q V G T Q K P E T H P P L A W K E C T S G C T S K D G S V V I T A N W R W H S V D G Y K N C Y T G N E M D S T L C P D D A T C A T N C A V D G A D Y A G T Y G A T T E D S L S V R F T G S N I G S R Q M F K U L N K E F T D V D V S T U P C S L N G A L Y F V S M D A D G M S K Y E T N K A G A K Y G T G C D S Q C P R D I K F P E G E N S E G E G G E D G N S C T G K Y G T C C A E M D I N E A N S T S T A L T P H P C D T M Q T R C S G D C T R D T G L C D A D G C D F P D G C D E N P F R M G N S F Y G P G K M D T K S R M T V T Q F I T S D G T D T G S L K E I K R V V Q N G K V I A N S D V S G I T G N S T S D F C T A Q Q K T F G D E D V T N K H G G S L S G M G D A L G E G M V L V M S L W D D H N S M W L D G E K Y P T D A A A S K A G V S R E T C T S D S G R E S T V E S E S S A K V V F N K V G S I G S T F S A

TABLE 7-continued

SEQ ID NO.	Amino acid sequence
SEQ ID NO : 85	MTSKIALASL FAAYGOOG TYTTEHPSL TWOSCTAKGS CTTOSGSIVL DGNRWRTHST TSSTNTCYTG TDATLCPDD ATCAONCALD GADYGTGIGI TTSGSLRIN FVTOANKV GSRYLLADN THYKPLDVS NPFIDVDS NPFCLNGAV YFANPADGG ISSINKAGQ YGTGCDSDOC PRDKFTNGK ANYDGWPS NNPNTCVGNY GSACAAEDMW EANSISTAVT PHSCDTVTOT VCTGMCGGT YSTRYAGTC DPDGCDFNPY RQENESFVGP GKTVDTMNSV TIVTOPLTD GTSSGTLINEI KRFVYQNGKV IPESESTISG VTMNLITPF CTAKTAFGD PTSFSDHGGIL ASMSAFAERAG MYLVLSLWDY YYANMMLWDY TYPTIKTAG GPRGTCSTSS GYPASTEASS PNAYVYVSNI KVGAINTFG
SEQ ID NO : 86	MYTKAALAA LVATVRGQAA CSITAETHPS LOWQKCTAPG SCYTTSVQGYT IDANWRHLQ TNISSTMCSS DTDCATKCCU DGADYGTGIG VTAWSNSLNK KEVTQGPYSK NIGSRWLMVE SESKYGQFTL LGQFTEFDV QD YFVSMDDL GGSKYTTNKG AGAKYGTGYC DSCOPRDIFKE INGOANIDGW QPSSNDANGA LGNIGSCCSE MDIWEEAKVY AAYTPHECTT IGOQMTGDD CGGTYSSDRY AGICUDGDDG FNSYRGMDSQ PRDKFTNGK GS KFTVTVTFQF TVTQPLTD GTSSGTLINEI KRFVYQNGKV IPESESTISG VTMNLITPF CTAKTAFGD DTNVEFEERGGD LAQMGKALAE PMYLVLSSWYD DHAVNMMLWDY STYPTDSTKP GAARGDCPFT SGYPADVEQSO APNSVVIYSN IRESPINSTY TGTPSQQNPP GGGTTTTT TTSPKSPGTQ TTNPSPGQQT HWGQCGQGW TGPTVQCSQY TOKSYNDWYS QCL
SEQ ID NO : 87	MYORALLFSA LLSVSRAAQO GPAQEYVPHS LTWQRCBASQ YCMEPLLINE FTEPDVDSNF PGGLNGALYF TSMDADGILS KYEENTAGAK DGDYFETAYG ITSENGDSLTL KEVTGSGNVS RVYLMEDDET YCMEPLLINE FTEPDVDSNF PGGLNGALYF TSMDADGILS KYEENTAGAK YGTGCDSDQC PRDKFTNGK GRVGEWBFSD SDANAGVGGM GTCCPBMIDW EANSISTAVT PHSCDTVTOT VCTGMCGGT YSTRYAGTC DPDGCDFNSY RMGNTTRFYGP GALIDTSSKQ TVTQPLTD GSLSIKRKY VONGEVINS ESNIQVEGN SITSEPECTAQ KTAFGDEDIF AQHGGLSAMG DAASAMVLL STWDHHISSM MWLDSYYPTD ADPSOPGVAR GTCEQGADP DVVSEHADA SVTFENIKFG PIGSTF
SEQ ID NO : 88	MMMKOYLOYL AAGSLMTGIV AGGGVGTQOT ETHPRLTWKR CTGKANCNTV QAEVVIDSNW RWIHTSGGTN CYDGNAWNTA ACSTATDCAS KCLMSEGAGNY QQTGCASTSG DSILTKEVTK HEYGTIVGSR FYLMGASKY OMFTLMNNEP TPDVDSLTV CGLNSALYFV AMEDGGMRS YPTNRAZAKY GTGYCDAQKA RDJLKFKVGEKA NIEGRVESSN DENAGYVPGY GCCEAEIDWVE SNAHAYAFTP HACENNNYHV CERDTCGTY SEDRFLAGGGK MNDDFCKG ANGCDYQKA YKDTDVKET VWTRODNLN QLAPVQNCQ ILAPSNLITE PCSTQFDVFTI DNRPREFVGD FQINNAIRI PMVLYNSWA DHTANMMLWDY STYPPKEKEA PGAARGPCAQ DSGVPESEVKA NYPNAKVVWS NTRFPFGST VNV
SEQ ID NO : 89	MYORALLFSP FLAAARAQQA GTVTAENHPS LTWOOCSSQG SCTTONGKV IDANWRWVHT TSQGYTCTG NTWDTSICPD DVTCAGNCAL DGDYSGTYG VTISGNVPS NEVTQSGKN DTTQIQLQD GOETPDLVDV SNLPGCLNGA LYPMAMDAD GLSKYPKGNKA GAKYGTGYC SOCPRLKFI NGANAVIGWV PSANDNAGV GNHGSCCAEN DVEANSTV AVTHPCTDP GOTMCODDCG GGTTSSTRYA GTCDDPGCDF NPFRQHNIFP YGPCKIVDTS SKFTVYQFI TDDGTPSGTL TEKRFYQVN GKVIQSEST ISGVTNSTI TEYTAQKA FGDNFTGFFTH GGQKIQSAL AJCMVLYNWL WDDHAANMLW LDSTYPTDAD PDTPGVARTG CPTTSQGPAD VESQENPSV IYSNIKVGP1 NSTFTAN
SEQ ID NO : 90	MFAIVLGLLT RSLGTGNOA ENHPSLWSQN CRSGGSCTQY SGSVYLDNN RWTHDSLSN CYDGNENSSS LPDPKTCSD NCLIDGADYS GTYCDAQCPH DIKPINGEAN VENWKPOUND ENAGHRYGA CCTENDIVEA NKVATAYPAH ITCVNGYRC DGSECDTDS GNRQGVCDK DCGCDNSYRM GNTSFWGQI ILDTGKPVGT DNGQSLSEIRR KYQGKVKTE VTMVNLAGMS SGNS1TDDFC NEQKSAGDT NDEPKGGGLS GLICKAFDGYM VLVLSLWDDH OVNMLWLSI YPTDQPSAOQ GVKRGPCTA SGAPSDEVSQ HPDSSVTFSD IRFGPFDSTY
SEQ ID NO : 91	MHORALLFSA LVGAVRAQQA GTLTEEVHP LTWOKCTADG SCTEOSGSIVY IDSNMRWHS TNGSTNTCYTG NTWDESICPD NEACAANCAL DGADYESTYQ ITTSGDALI TTVTGENVGS SYQFEDLVN EFIDFIVSN LPGCLNGLY FTSMADGGV SKYPAKNGA KYGTGCDSDQ CPROLKFING MANEGWTPS DNDKNAVGGG HGSCCPEDLN WEANSSIAF TPHCPDDLGQ TMCSGDCCG YSTRYAGTC CDPDGDFENA YRGMNTSYG PKIVDTSV MIVVTOPIGD GGSLSEIKRL YVONGKVKIAN AQSNDVGTG NSITSDECTA QKTAFGQDI FSKHGGLSGM GDAMSAMVLL LSIWDDINS MWLDSYTPD DADSEGPVIA RGTCHEHGPQ PETVEQHGPV AVTFPSKIF GPIGSTYSSN STA
SEQ ID NO : 92	MFRAAALLAP TCLAWSQO AGTNTAENHP QLOSOQCTS GGCKPLSTKV VLDSNMRWHS STSGYNTCYTG GNEMDTSLCP DGKTCANCA LDGADYSGTY GTRSTGTAI LQFVTLGSV SRVYLMADDT HYQLKLRLQ EFIDFIVSN LPGCLNGLY LSAMPDAGM SKYPAKNGA KYGTGCDSDQ CPDKIKEING EANVGNWET GSNTGNGSYG TCCSEMDINE ANNDAAPTP HPCTTQTR CSGDIDCARNT GLCIGDGCDF NSFRNGDKEF LGKRMVTDIS TNDNTSTGQF TPHCPDDLGQ TMCSGDCCG YSTRYAGTC FSKHGGLSGM GDAMSAMVLL LSIWDDINS MWLDSYTPD DADSEGPVIA RGTCHEHGPQ PETVEQHGPV AVTFPSKIF GPIGSTYSSN STA

TABLE 7-continued

SEQ ID NO.	Amino acid sequence
SEQ ID NO : 93	MRTALALILA LAAPSAVSAQO QACTITAEFH PTLTIOQCTQ SGCCAPLTIK VVLDVNPWLI HSTTGTYNCY SGNTDAILC PDPVTCACNC ALDGADYTGT FGILPSPGSIV RLFLIAIDSH YOMFPLINKE FTFDVEMLNNE RGCGSAGHL TAMDADGLA KYPENOAGAK YGTGKPSAQ PKGKFINGQ ANVEGTFPQ CTDIALWE NDNSAFAH PCTTNSQTRC SGSDGCNFN LCDADGTLF DDLCPQENAA FGGTSFAOH SERMENTTFF GAGMSVDTK LFTVWTGTOIT SDNTSNGALV EIHRLYUNQ QVIONSYVNI PGUNPATSIT DDLCQENAA AGSALFSGRS PGPVPGSAP ASSATATAAPP GLAQGEAL RSGNVLASI VNSAADTFLW DSNYPADAP SAPGVARGTC PDSASATPEA PTPSVFSNI KLGDGTTFG AGSALFSGRS FGSQCGLGIL AGPTGVCPSP YTCQALNIVY SQCI
SEQ ID NO : 94	MYORAUFLSPF ELAAARAHFA CTYTAENHPS LTWQOQSSGG SCUTONGKVU IDANWRWYHT TSGTYNTCYTG NTWDTSICPD DVTCQAONCAL DGADESYTG VITSGNAILR NEVTQSSCRN IGSRLYLLQ DTTIQIFKLQ GQETFTDVV SNLPGCLINGA LYFVAMDAVG NLSKYPGNKA GAKYESTGYCD SOCPRDLIKEI NGOANVGSWQ PSANDENAVG GNHGESSCAEM DYMENASTST AVTPHPCDTP GOTMEOQDDC GTTYSSTRYA GTCDCDGDGF NYPQPGNHSF YGPQKIVDTS SKFTVVTQFI TDDGTPSGTL TEIKRFYQN GKVIQPSBT ISGVTVNIT TEYTAQKAA FDNTGFRTHG GLQKISQALI QGMVLVNSLW DDHAANMLW DSTYPTADP DTFGVARIGC PTTSQGPADV ESQNPNSYI YSNLKVGIPIN STTIAN
SEQ ID NO : 95	MHKRATLSA LUYYAAGFAR GGCGTQOTE THPKLTFRQK SAAGSCTTON GEVYDIAWLR WYHDKNGYTN CYTGENEMNTI ICALAAASCAS NCWVGDADYO GLTGAISTSGR MDEOGTTSKY PSNKAGAKY TGYCDSCOPR DLKFDGDKAN SYASWQPSSND ONAYGIGMGWS CCHEMDTNEA NSVSAATPH PCQNYQHSC SGDGGTYS ATTRAGDCDP DGCDDNATRM GHDFYNGK TVDTGKFKSI VTQFRKGSGST LTELKOFVQ DORKINPNA TWPGLBPENS ITPDFCKAQK QVFGQPDREN DMGEGTNNAK ALANPMVLV SLWDHYSNM LWLDSTYPTD ADSSAPGKGR GTCDTSSGP SDVESENGDA TVISNIKFG PLDSTYTA
SEQ ID NO : 96	MRSALLAFSL NSAAGQQAGT LQTKNHPSLT SQKCROGGCP QVNTTIVLDA NWNRWTHATSG STNCYGTNTW QATLICPDGKT CAANCALDGA DYTGTGIVTT SGNSLTLOFV TOENVGARLG YLMADDTYYQ MFMLNQEFW FYDMSNLPC GLNGALFVSA MARTAWMPM VVCATPLIS TRRSTARLLR LPPVPPRSYK RGICDSCQCP DIKFINGEAN VQKQPSND TNAGTGYKA CCONKMDWEA NSISATAYTPH PCTQRGIVRC SGTA CGGGSN RYGSICDHD LFQNOLRNG RTRSVARVGR VKOFNRSRVS VEPIHNLWPK SDCHVQNGR VIONSKVNTI GMPSMTDVSU TERPAQHSTI FNDTFSPOQ CGMANMSEI PRCMVLISI WDHANMLW LDSDHANMLW RSTPSEHAT PIRSEQRSS HSTQTRYVTP TMKEGDPNS TGTTYTGGSV PTSTSTGTT GSTPQQTGTG VTVPOQOCGG IGYTQPTCA SPITCHLNP YYSQCY
SEQ ID NO : 97	MKOYLOLAA ALPLMSLYVA OGYGTSTSET HPKITWKKC SGGSCSTVNA EVVLDVNPWLI LHNADSKNCY DGNEWTDACT SSDDCTSKCV LEGAETYKTY GASTGDSLIS LKPLKTRFLY TMNGSKYOMFT LMNNERADV DLSITPCVGLN SALYFVAMEE DGMNASYSTN KAGAKYGTG CDACCARDIK FVGGKANYDG WTPSSNDANA GVGAIGGCCA EIDWENSAH AFAFTPHACE NNHYVCEDT TCGTYSEDR FAQGDANGC DYNPRVGET DFYKGKMTD TSKKPFIVSQ QENKLTOFF VONGKKEIP GPKHGELEPTE SSDIPELCS AMPEVFGDRD RFAEYGGFDA LNKAALAVPNV LYMSIWDDH ANMLMLDSY PPEKAGTGG DRGPQADSG VPSEVEQYP DATVWSNIR FGPGSTVQV
SEQ ID NO : 98	MFPKASLIAL SFTAVYVGQO VGTOMAEVHP KLPSONLCTS GCNTNONTAVV LDANWRWHLT TSGTYNTCYTG NSWDATLCPD ATTCAQNCAV DGADYSCTYG ITTSIGNALTL KFKTGTVNGS RVYLMQDFTA YQMLQNLNE PTTFDVMNL PCGLNGALYL SQMDQDGGLS KFPPTKAGAK YGTGKCDSOC PHD1KFINGM ANVAGWAGSA SDPNAGSTL GTCCSEMIDN EANNDAAFT PHPCSDVGDOT QCSGFCGDD DERSGLCDK DGCDFNSFRM GDKSFLGKWM VTFQVFTTFTV DTSRKEFTV VTFQVFTTFTV VTFQVFTTFTV VTFQVFTTFTV VTFQVFTTFTV VTFQVFTTFTV NYFATLNTFE KMGRALKSOM VILAMSVNDH AASMONLDSN YPADDDATK GVARGTSEAD SGLPNTYESQ SASASVTFSN IKWDINTTF TGTGTSPPS PAGEVSSSTS VASOPTQPAQ GTVAQNGQCG GTGFTGPTYC ASPPTCHVN PYHSCY
SEQ ID NO : 99	MFRTAALLSF AYLAVYVGQO AGTSTAETH PLTWEQCTG GSCTQFQSSV VLDNMRWTH VYGGTYNCY GNEWNTVCP DGTCAANCA LDGADYEGTY G1S1SGNALT LKFVTAQOT NVGSRYLMA PGSETEYOMF NPNUOEFTD VDVSALPCGL NGALPSEMD ADGSESEYPT NRAGAKYGTG YCD5QCPDI KFIEGRANVE GWTPSSTS PN AGTGTGICC NEDIWAEANS ISEALTHPC TAQGFTACTIG DSCSPNSTA GLCDQAGCOP NSPFGDTSF YCPGLTVDT SKITYVTOFI TSDNTTGTDL TARRIYTON GOVIONMSN IAGVPTNEI TTDRDQOKT AFGDNTNFESE KGGJTGMAZIA FSGMGMVLIS IMDDDAEML WLDSTYPTVKG TGPAAARTIC ATISQGPDDQV ETQSPNAQVV FSNIKFGAIG STESTGTGT GGTGNGTGT GUTTSSAPA TQTKYQCGG QGMIGATVCA SGSTCTSSGP YYSQCL
SEQ ID NO : 100	MERTAALTAP TEAAVVLGQO VGTLTENHP ALSIQCTAT GCTTQOKSYV LDNNMRWTH TAGATNCYTG NAWDPALCPD PATCATNCAL DGADYSCTYG ITTSIGNALI RFTVNGQYQSO NIGSRVYLL DADHKLFDL KNOEFTFDV DMSGLPGSLNG ALYFSEMAAD GGKAHAHAGNN AGAKYGTG YCDSQCPDI KFIEGRANVE GWTPSSTS PN AGTGTGICC NEDIWAEANS ISEALTHPC TAQGFTACTIG DSCSPNSTA VCDKDQGDFN SYRQGDKNFL GRKTIDTJK KVTVTQFIT DNNTPTGNLVY EIRRVYVONG VYQNSNFTF PSLSQNNTSIS DEFVQAQKTL

TABLE 7-continued

SEQ ID NO.	Amino acid sequence
SEQ ID NO : 101	FGDNQYNNH GGTTKMGDAF DNGMVLIMSL WSDHAHMLW LDSDVPLDKS PSEPVGSRGA CPTSSDDDD VVANHPNASV TFSNIKYKPI GSTFEGSTPP VSSGGSSVPP VTSIISSGT TPTCPTGTVK KNGQCGGKGY SGPTACVAGS TCITSNDWYS QCL CSLTPETKPA LSWSKCTSG CSNTQGSTVTI DANWRTHOL SGSTMCYTGK KWDTSCTSG KVCAEKCCID FVTKGTGTVN IGSRTVLMED ENTYQMPQLL GNEFTFDYDV SNIGCGLINGA LYFVSMDADG GKAYPGNKA GAKYCTGYCD AOCPRDVFELI NGOANSQDMW PSKSVDYNGI GNLTGTCPEM DIWEANSTIST AHTPHPCITKL TOHSCTGDSC GGTSEDRYG GTCTCADCDF NAFQGNKDF TIKKVTVMLI FHKGSNRLLS EITRGLVSGKVI KVANSEKSI AVPNVSLPFT EFCATQKKVF GDIDDEFEKKF AWGMSDIAFL DSTYPTDSTK LQGARQGKST SSGVPADLK NVPNSKVAWS NIKRGPIGST YKEGOEPTIN PTINPNTIPG GTVDQWQCGK GTNYSQPTAC KSPFTCKKIN DFTSQCG
SEQ ID NO : 102	MFRITALLAF TMAAMVFGQQ VGTNTAENDH TLTSQKCTKS HGCMSNLNTKII VLDANWRWLH STSGYTMCY GNWIDATLCP DGKTCANA CA LDGADYGTGY GITASGSSUK LQFVLTGSGNV SRVYLNMADDT HYQNBQLNQ EIFFEDVDMN LPCLGILGALY LSAMIDAGGM AKYPTNKGKA RKGTYCDSQ CPDIKIFING EANVEGNNA SANAGTGNYG TCCTEMDIVE ANNDAATIP HPCTTNAQTR CSGSDCTRTD GLCIADGDFP NSFRNGDQTF LGKLGTVDTS KFTVWVQFI TNDGTSAGTL TEIRLYVQN GKVNQNSVKV IPGIDLVNSI TDNECSQKQT ARGINTNYFAQ HGGLHQVGEA LRTOVNLALS IMDYAANML WLDNSNTPNK DPSTPQVARG TCATTSGVPA QTEAQSPNAY VVFSPNTRKFGD LATNTYTGTYS SSVSSSHSS TSTSSSHSS
SEQ ID NO : 103	MYQTSULASL SFLATSAQO QVGTQTAETH PKLTQKCTT AGGCTDQSTS IVLDANWRWL HTVDGTYNCY TGQENDTSIC TDGKTCAEKC ALDGADYEST YGISTSQDNL TNIKGRVYL EDNEYDQMFEN LINKEFTFDV DVSVKPGCNG LNGALYFSEM DSDGELSKYT TNIKAGAKYGT VGEVAENHP RLPMQETRN GGCTQVSNQ VVLDANWRWL HTVDGTYNCY TGNANNSVC SDGATCAQRC NRYGEVCKDKD CDSRCPMDK FVCPGKTVT WNKMTVUTQH LDIRLRLVYQ SGRVYAPCQ TVNDGTYPAHD QGRDCNQPPY AAQKTLSEDII NDFTKKGIA RMGESPERGM VLVMSVWDDH AANALWLDSS YPTDADASKP GVYKGPCTS SGVPDVEAN DADSSVIYSN TRYGDIGSTF NKTA
SEQ ID NO : 104	MFSKVALTAL CFLAVAAQAO VGEVAENHP RLPMQETRN GGCTQVSNQ VVLDANWRWL HTVDGTYNCY TGNANNSVC SDGATCAQRC ALLEGANYQT YGISTSQDNL TNIKGRVYL EDNEYDQMFEN LINKEFTFDV DVSVKPGCNG LNGALYFSEM DSDGELSKYT RAGAKYGTGY CDSRCPMDK FVCPGKTVT WNKMTVUTQH LDIRLRLVYQ SGRVYAPCQ TVNDGTYPAHD QGRDCNQPPY EGLCDPQGCD YNPPRMGKDF KFTVWVQFI TNDGTSAGTL GSIALTSELP HPCTVQQTW CSGDSCEGTY SNDRGYGTCD KRAFDNDSP SFTGQKTVL RSLAKGHLA LSIVNDHTIA MLWLDNSNPT DAPPNKPGIA RGTCPVTTGS PRDTEQHPD AQVIFSNIRE GDIGSTFGN
SEQ ID NO : 105	MYRKLAVisA FLAAARAQV CTQQAETHP LTWOKCTASG CTPQOGSVVL DANWRWHTD KSTTMCYDGN TWSSFLCPDD ATCAKNCLD GANYGTYGV TTISDALTIQ FVTAISNGSR LYLMANDSTY QEFUUSNFER SFYUDVSQFL CGLNGALYFV SMDAGQSK YPNAGAKY GTGTCDSQCP RDLKFINGQAI NYEGWEPSSN NANTGYGGH SCCEMIDIVE ANS1SEALP HPCTVQQTW CSGDSCEGTY SNDRGYGTCD PDGCDWNPYR LENTISFYCPG SSFALDTKPK LTIVTQFATD GSISYYVQN GVFOOPHAO VGSYSENTIN TDYCAEOTA FGGTSFTDKG GLAQINKAQ GGWLVMSLW DDYAVNMLW DSTYPTCATA STGAKRQVQ STSSGVPQV ERQSPNSKVI YSNIRFPQG STGGNTGSNP PGTSITRAPP SSTSSSPAT OTHYGOCCG GWTGPTRCAS GYTCQVLPNF YSCPL
SEQ ID NO : 106	MRASLLAFLS AAAVAGGCGA GTLTAKRHPs LTWOKCTRGG CPTLNNTMVL DANWRWHTAT SGSTMCYTGK KWQATLCPDG KSCAANCALD GADYGTGYI TGSEWSLTLQ FVTDVNGARA YLMADDTQY MLELINSNIPC GLNGALYFV MDAGWAKY PTNAGAKTA TGYCDAQCPR DLKXINGIAN VEGWTPSNTD ANGIDGHGSC CSENIDIMEAN KVSTAFTHP CITIEQNCIE GDSCGTYSD DRYGTLCDAD GCFDENSYRMG NTIFYGEKTI VDTISKFTVQ TQFIDSAGD LAETKAFYVQ NGKVIENSQS NVGVAEINSI TQSFKTSQKT ARGIDDDFNK KGKHKOMGKA LAQAMLYMS IMDDHAAANML WLDSTYPTVPK VPGAYRGSGP TTSGVPAEVD ANAPNSKVAF SNIKFGHLGI SPFSGGSSGT PPNSPSSSSA PTSSTAKPS TSTAANPSGT GAHHWAQGG IGFSQPTCP EPTCAKHD IYSQCV
SEQ ID NO : 107	MLASTPSYRM YKIALILAAI LGSGQAQVG TSOAEVHPSM TWQSTAGGS CTTNGKGVVI DANWRWHTV KDYTMCTGN TWDTKTLCPDD ATCA-SNCALAE ANTCA-SNCALAE GANYOSTYGA TTSGDSLRLN FVTTSOQKNI GSRLYMMKD TYEMEKLUN QFETEDYDVS NLPCGNGAL YEVANDADGG MSKYPINKAG AKYGTGCGDS QCPRLKPIN QGANYBGWQF SSNDANAGTG NHSSCAEMD IWEANSLISTA FTHPHCDTPG QVMCTGDAGC GTYSSDRYGG TCPDPGCDNF SFRQGNTTFFY GPGMTDFTKS KFTVYUQF IT DDCTAGSLK ELKRFVYONG KVIPUSESTW SGVGENSITN DYCTAQKSLF KDQMVFAKHG GMGNGMGAIA QGMVLMWSL DDHANMLWL DSNTPTASS STPGVARTC DISSGVAPDVA EAHIDDASTVV YSNIRKVGPIG STENSGGSNP GGGTTTAKP TTITITAGSP GGTGTAQHYG QCGGNGWQGP TTICASPTTCQ KLNDPFSQCL

TABLE 7-continued

SEQ ID NO.	Amino acid sequence
SEQ ID NO : 108	WQIKOYLOYL AAALPLVNMIA AORAGTQOT ETHEPLSWKR CSGGNCOTY NAEIVIDANW RWHLDNSYON CYDHEWTSA CSSATDCAK CYLEGANITY TIGVTSQSDA LTLKFKVTHI YGTNIGSKY LMNGSDKYM FLMNNEPAP DVDSLKVBCG LNSALVPSAM BEDGMRSTYS SNKAGAKYGT GYCAQCARD LKFVGGTANI EGWRPNTNDI NAGTPGYAC CAEDWWSIN AVAFATPHG CLNNYHVC E TSNCGTYSE DRFGGLCDAN GCDINPFRMG VDTSRKPTVV TRFEENKLQF FFQDGKID IPEPTWGLP NSSATPELC TNLSKVFDDE DRYEETGGFR TINEALRIPM VLWMSLWDHV YASMLWLDHV YPPEFAQGP AERGPACAPS GYPAEVQAQF PNAQYTWSNI RFPGPGSTYQ V
SEQ ID NO : 109	WTSRITALVSL FAAVYQGOYG TYQTEPHSL TWOSETAKGS CTNTNGSIVL DGNWRWTHGV GTSTMCYTG TMWATLCPDD ATCAQONCALE GADYGTGYGI TTSNSLRLRN FVQTSANKNI GSRYVLMADT THYCFNLIN QEFITFDVDS NULPGUNGAV YFAMLPADGG ISSTNTAGAE YGTGTCDSQC PRMKFTRGO ANVGDWYPS NNANTCVGNH GSACCAEMDIW EANSISTAVT PHSCDVTOV VCTGDGCGT YSSRYAGTC DPDGCDENSY RMGBDBTFVGP GKTVDTMSVF VTVTQFLTD GTASGTLNEI KRFVQDGKV IPNSYSTISG VSGNLTTPF CDQKTAFGD PTSFSDDHGGL ASMSAAFFAG MYLVLSLWMD YYANMLWLD SVTYPGKTAGS GPRGTCDSI GYPAVEAASS PNAYVYYSNI KVGAINTYGS
SEQ ID NO : 110	MFEVYLWLT OSLGTTGIIQ ENHPHSWQN CRSGGSCOTQ SGSTVLDNSW RWTHDSLTN CYDGMENSS LPDPHTCSD NCLIDGADYS TGYGTCQSPH DLEPFGHEAN PYTSNLSLRN SLKLWVFTING PTKLNEKTFI FTYKQDNEP MDEGOTTSRF SSNAGAKYTG DGCDAQCDSC PRDLIKEPQ ANVEGWIGS TDNSNGTGYN GTCCSEMDIW FANSAAYT PHPCSTNQQT RCTGADCGQ DDRDGVCDFP DGCDDNSFRM GDQFLGKOL TVDTSFIDDTG DNGQUSEIRR KTYQGKVKIE NTUVNNTAGMS SGNSSTDFFC NEQKSAGDGT NDPEEKGGGLS GLGAFADYGM VLVLSLWMDH QVNMLWLDSTI YPTQDQASQP GVREGPACATS SGAPSDEVSQ HPDSVSTFSD IRPCEPIDSY
SEQ ID NO : 111	MERKAALLAP SELIAIAHQGQ VGTNQAEQH SLPSQKCTAS GCTTSSTSYY LDANWRWHT TTGTYTCTG QTWDASICPD GYTCAKACAL DGADYSGTGY TG ITTSQNALTI QFVKGTTNGS RVYLLDASN YOMFOLINQE FTEDVDMENL PCGLNGAVYL SQMDQGVS RFPNTAGAK YGTGTCDSQC PRDLIKEPQ ANVEGWIGS TDNSNGTGYN GTCCSEMDIW FANSAAYT PHPCSTNQQT RCTGADCGQ DDRDGVCDFP DGCDNSFRM GDQFLGKOL TVDTSFIDDTG DNGQUSEIRR KTYQGKVKIE NTUVNNTAGMS SGNSSTDFFC TOQKTAFGDGT NPEAAQGKQ OMCAZLKSQW VLAJSLWMDH QVNMLWLDSTI YPTQDQASQP GVREGPACATS SGAPSDEVSQ HPDSVSTFSD IRPCEPIDSY TGTLLTTPSGS SPPSPASTIS GSSTASSASA SVPTQSGTVA QWAQCGGIGY SGATTCVSPY TCHVNNAYYS QCY
SEQ ID NO : 112	MYRAJATASA LIAARAQOQ CTLTETKPA LTWSKCTSG CTDVKGSTGI DANWRWHTQ SSSTMCYTG KWDTSVCTSG ETCAQKCCLD GADYAGTGYCI TSSGNQLSIQ FVTKGSLTMEN IGSRTILMEN ENTQOMQLL GMEFTDVDS SNIGCLINGA LYPSMIDAG GKARYPANKA GAKYGTGYCI AOPCRDVKP NGRANSQWPK PSDSDINAGI GNMGTCPPCEN DMEANSTT AFTPHPTKTL TOHATGDS C GTCDADGCCDF NSTROGNKTP YGRGSDENID TTKKTVTVTQ RKKGSNGLRS ETPLVYQONG KYTANSEESKI PGNSGSSLLTA DECSKQKSYF GDIDDRFSKKG GWSONSDALE SPMPVLMNL WHDHHSNMLW LDSTTPDST KLGQQRGSCA TISGVPSDL RDVPNSKVSF SNIKFGPIGS TYSSTGTTNPP PSSIDTSTTP TNPTGGTVG QYQQCGQGY TGPDKCKSPY TCKKINDFYS QQ
SEQ ID NO : 113	MSSPQIYRAA LLLSLILATAN AQQVGTITE THPSLITWQTC TSDGSCTTND GEVIDANWR WHSTSEATN CYTGNEWDTI ICTDDVTCAA NCALDGATYE ATYGVTTGSSC SSKNITQPSL LMSDNTSYN FLKLGQFET DVIDSNLUPCG LNGALYFVAM DADGCTSEYS GAKAAGAKYGT GYCDSQPRD LKPFINGEAN DCWEPSSNVN NGVYGDHGS CAEMDWWEAN SLSNMFATPH GDSCCGTYSQ SGDRTSGTCD PDGCDYNPYR LQNTDFGPQ LTVDNTISPFT VTOCTSLTEIK RLYVQONEBI ANGASTYSSV NGSSUTSAFC ESEKTLFGDE INPDKHGGL E MGEEAMAKGM VLVLSLWMDY ADMWMLDSD YPNASSATP GVARGTSTD SGVPATEAE SPNAVVTYSN TKFGTGSTY SSGSSSGSCTT ATSTLTKTS TTSSSSSTS AAQAYGQCGQ QWMTGPITCV SGYCTYENA YYSQCL
SEQ ID NO : 114	MHORALLFSA LITAVRAQOA GTLTEEYHPS LTWOKCTSEG SCTEOSGSVY IDSNWRWHTS VNDSTMCYTG NTWDAFLCPD DETCAANCAL DGADYESTYQ VTIDGDSLTL KFTVGSANGS RLYLMDTSDE GYOFNLNDA EFTEDVDSN LPCGLNIALY PTMADGGV SKYPAKAGA KYGTGTCDSQ CPDLKFKTDG QANVDGMEPS SNNDNTGIGN HGSCCPENDI WEANKLSTAL THPHCOSSEQ TMCEENDCGG TYSDRYGST CDPDGCDENP YRMGNDSFVQ PGKTDITGSK MTVVTQFIDT GSGSLUEIR YVQNGVIA NADSNTSGVT GNSISITDFCT AKKAFGDED TFAEERINGLAG ISDAMSSMVL ILLSLMDIYYA SMEWLSDYP ENATADPSV ARCTCDSSEG VPATVGEAHP DSSVTFENIK EGPTNSTFSA SA
SEQ ID NO : 115	MYAKPATLAA LVAGAAQNA CTTAENHPS LTWSKCTSG SCTSVQGSIT IDANWRWHTR TDSATHMCYEG NKWDTSYCSD GPSGASKCCI DGADYSSSTYQ ITTSQNSLNL KFTVKGQFST NIGSRTILMNE SDTKYQMFQL LGMEFTDVDS ALFYFSDAD GMMSKYSGK AGAKYGTGYC DSQCPDLKFE INGEANEVNW QSSTNDANAG TGKYGSCSCSE MDWMEANINA AAFTPHPCXV IGOQSCEGDS CGGTYSTDRY AGICDPPGCD FNSYRQGKNT FVYKGKMTVDT TRKKITVVTQF LKNSAGELSE IKFYVQNGK VIPNSESTIP GVEGNSITQD WCDQKAARG

TABLE 7-continued

SEQ ID NO.	Amino acid sequence
SEQ ID NO : 116	MVAKPATLAA LVAGASAQAV MVONGKALAG PMVLVMSIWD DHAVMMLWLD STWLDPGAK PGAEERGACTP TSGVPAAVEA EAPNSNVIFS NIRFGPIGST VSGLPLDGSSG NNPPVSSST PVPSSSTTSS GSSCPTGCTG VAKHYEQCGG IGFETGTQCE SPYTCNLND WYSQCL CSLTAETHPS LTWOKCTAPG SCNTVAGSIT IDANWRTHO TSSATNCYSG SKWDDSSCTT GTDCASKCCI KFTVKTGQST NIGSRVLYME SDTKYQMFKL LGNEFTFVD VSNLGCLNG ALFYSMAD GMMSKYSGNK AGAKYTGTGCY DAQCPDRIKE INGEANTEGW ESSSTDANAG SGKYGSCCTE MDWEMANNMA TAFTPPICTT I GOTREGETD CGGYTSSDRY AGVCDEDGCD FNSRQGKTVT DKAQIIVDQF LKNSAZELQVTP VENSESTVKA YCDQKTVFO NTDDDETAKGK DVYLVMSVWDF DHAVMMLWLD STYPTDQGVY AGERGACTP TSGVPAAVEA NAPNSNVIFS NIRGPPIGST VQGLPSSGGT SSSSSAAPQS ISTKASTTS AVRTSTATT KTSSAPAAG TINTAKHMQC GENGWITQPTV CESPYRKTQ NDWYQSCL
SEQ ID NO : 117	MLTLYVFLLS LVVSLIEGTQ QSEDHPKLTW QNGSSSVGS IVLDNSNMRWV HDSGTTINCYD GNLSKDLCP SSDTCSQKCY LEGADYSGTY GLQSSGSKLT LKEVTKGSEY TNIGSRVYL KDENTYESTFK LKNEKFPTVY DDESKLNCGUN GALYFVAMDA DGGKAKYSSP KPGAKYGMGZ CDAQCPHDIM FISCKANVDD WKPQDNIDNS GNGKLGTC S EMDIWEGNMK SQAYTVHACT KSGQYECTQ QCGDTSGDR FGKTCDKDGC DYASPRWRGDQ SFYEGEKTVIY PIGDPLTEIR RLYVGQKTI NNEKTSNLAD TYDSITDKFC DATKESASDT NDFTAKGAMS GFSTNLNNGQ VLYMSLWDDH TANMMLWLDST YPTDSSSTA QRGPCTPSSG VPKDVESEORG DATVVFSDIK FGAINSTFKV N
SEQ ID NO : 118	MLAAALFTFA CSVEVGITKTP EDHPKLUWQN CASKGCSQV SGEYTMDSNW RWTHDGNGRN CYDGNMTWIS LCDPDTICSD KCVLDGAETYQ ATYGIQOSNGT ALTLKFEVHG SYSTNIGSLR YLLDKSTYY VEKLNNKEFT FSYDVSKLPC GLNGALYFVE MDADGKAKY AGARPGAETYQ LGYCDAAQPS DLKFPINGEAN MGDKTFEPGP LIIDTKSPEV VUTQFQYSPV TEIRKQYVN GKVIENNSKN I PGDTCQKKA ARGDTNDFKN RGGFPAKLQVW FDGRMVLYLW LDHDHQVAML WLDSTYPTNK DKSQPGVDRG PCPTSSGKPD DVEASASADAT VVYGNIKFGA LDSTY
SEQ ID NO : 119	MLTLYVFLLS LVVSLIEGTQ QSEDHPKLTW QNGSSSVGS IVLDNSNMRWV HDSGTTINCYD GNLSKDLCP SSNTCSQKCY LEGADYSGTY GLQSSGSKLT LKEVTKGSEY TNIGSRVYL KDENTYESTFK LKNEKFPTVY DDESKLNCGUN GALYFVAMDA DGGKAKYSSP KPGAKYGMGZ CDAQCPHDIM FISCKANVDD WKPQDNIDNS GNGKLGTC S EMDIWEGNMK SQAYTVHACT KSGQYECTQ QCGDTSGDR FGKTCDKDGC DYASPRWRGDQ SFYEGEKTVIY PIGDPLTEIR RLYVGQKTI NNEKTSNLAD TYDSITDKFC DATKESASDT NDFTAKGAMS TANMMLWLDST YPTDSTKTAQ SRGPACVSSG VPKDVESQYQ DATVYTSDFIK FGAINSTFKW N
SEQ ID NO : 120	MILALLSLAK SLGATNQAE THPKLTWTRY QSKGSGOTVN GEIVDLSNVR WTHHSGNTNY DGNTWSLSC DPPTCSNNC DLGADYDPTG YGISTGSNLSL KLGFTVTHSY STNIGSRVYL LRDSKNEYMF KLKNKEFTFI VDDSKLPCL NGALYFVAMD EDGGSKNSI NKAGAQYGTG YCDACOPHDIM KFTINGEANYL DMKPOSDEN SGNGRYGACC TEMDIWEANS MATAYPHVC TVTGLRICEG TECGTDANO RYNGICDKDQG CDENSYRLGD KTFPGVGTWV DSSKPVTVT QFVTENGQDS GTLSBIRKY VQGGKVLENS KVNLGIAAG NSVTDTCNE QKKAFGDDND FEKKGGHLKGIG VLSLWDDHSV MLWLDSTYPT DSGSKFAGADR GPCATSSGVP KDVESNTYASA SVTFSDIKFG PIDSTY
SEQ ID NO : 121	MVIALLILRG LSVTGTQOOS EHPKLTWTRY QSKGSGOTVN GEIVDLSNVR WTHHSGNTNY DGNTWSLSC DPPTCSNNC DLGADYDPTG SGTGTTGSGV SYLGSITNGS VYLLGDSB TSLLKLGFTV TFPVDDNEF KFLKLENNEF VDSSKLPCG NGALYFVAMD AMEDGASK YSGAKPGAKY GNGYCDAAQCP HDKWFINGD A NYEGKPSDN DENAGTKGWY ACPEMDINE ANKUTATYTF HICKNGEYR CEGTQGDTK DNNYQYGVCD KDGCDENSWR MGROSFWGPG LIIDTGPFTV VVTQFLADGG SLSERPKVY QGGKVIENTV TKISGMDBFD SITDECRNQ QKARPDNDFE EKKGGHLKGIG TAVIDGVLV LSLWDDHSV NMWLDSTYPT TNAAGAAGLT ERGACATSSG APSVUEQSP DATVTEFDIK FGPDSTY
SEQ ID NO : 122	MLLALFARGK SLGATNQAE NHPKLTWTRY QSKGSGOTVN GEIVDLSNVR WTHHSGNTNY DGNTWSLSC DPPTCSNNC DLGADYDPTG YGISSGNSL KLGFTVTHSY STNIGSRVYL LRDSKNEYMF KLKNKEFTFI VDDSKLPCL NGALYFVAMD EDGGSKNSI NKAGAQYGTG YCDACOPHDIM KFTINGEANYL DMKPOSDEN SGNGRYGACC TEMDIWEANS MATAYPHVC TVTGLRICEG TECGTDANO RYNGICDKDQG CDENSYRLGD KSFPGVGTWV DSSKPVTVT QFVTENGQDS GTLSBIRKY VQGGKVLENS KVNLGIAAG NSVTDTCNE QKKAFGDDND FEKKGGHLKGIG TNAAGAAGLT ERGACATSSG APSVUEQSP DATVTEFDIK FGPDSTY
SEQ ID NO : 123	MLASVYLYS LVVSLIEGTQ QSEEHPKLTW QNGSSSVGS IVLDNSNMRWV HDSGTTINCYD GNLSKDLCP NADTCSKCY LEGADYSGTY GITSSGSKVY LKEVTKGSEY TNIGSRVYL KDENTYESTFK LKNEKFPTVY DDESKLNCGUN GALYFVAMDA DGGKAKYSSP KPGAKYGMGZ CDAQCPHDIM FISCKANVDD WKPQDNIDNS GNGKLGTC S EMDIWEGNMK SQAYTVHACT KSGQYECTQ QCGDTSGDR FGKTCDKDGC DYASPRWRGDQ SFYEGEKTVIY PIGDPLTEIR RLYVGQKTI NNEKTSNLAD TYDSITDKFC DATKESASDT NDFTAKGAMS GFSTNLNNTAQ VLYSVHCGMI IQPICCGHLR RIQRIOQKV QAVDRLCRR VFQRMKLSM VMLQSRTRTL SLELSSTRPLV GISBAGRLEF F

TABLE 7-continued

SEQ ID NO.	Amino acid sequence
SEQ ID NO : 124	MILALLVLGK SLGIATNQAE THPLKLWTRY QSKGSSTVN GEIVLDNSVR WTHHSGTINCY DNTWTSCLC DPPTCSNNC DLDDADYPTG KTGISTSGNSL KLGIVTYSY STNIGSYVYL LRDKNYEMF KLKNKEFTFT VDOSKLPCGL NGALYFYAMD EDGGSKNSI NKAGAQYGTG YCDACOPHDW KFINGEANVL DMKPOSDEN SGNGRGTACC TEMDWEANS MATAYPHVC TVTGLYRCEG TECGDNDQ RYNGICDKDG CDENSTRLGK DSSKPVTVTQ QFVTSGQDS GTLSEIRKRY VQGKVIENS KVNVAGITAG NSVTDTCNE QKKAFGDNND FEKKGGLGAL NMWLDDHSTP TNAAGAALGT ERGACATSSG KPSDVESQSP DATVTFSDIK FGPDIDSTY
SEQ ID NO : 125	MLCIGLISFY YSLGVGNTTA ETHPLKLWTKN GGQTGYNGEVT VDSNWRWHTH KGSTKNCYDG NLWSKDLCPD AATCGKNCVLU EGADYSGYG KTGISTSGNSL KFTHGTYSTV NGSRLYLK DEKTYOMFLN NGKEPTFTVD VSNLPGCLNG ALHYHNMED GGTRKVPDNE AGAKYGTGVC DAOCPDHLK INGPNSDGM KPOSNDKNSG NGKYGSCCSE MDIWEANSIC SAVTPHVCNDN LQOTRCQTA CGENGGSRF GSSCDPDGCD FNSWGNKNT FYGPGLIYDTH TYNMLWLDHV YPTNSKKGAS DRGPQPTSG VPADVEKSAA DANVYSDIR FGAIIDSTY
SEQ ID NO : 126	MLGALVALAS CIGYGNTMPE KHPDLKWTNG GSSVSGSIVV DSNWRWTHK GETKNCYGN LMSDKYCPDA ATCGKNCVLE GADYSGYGV TTSGDAATLK FVTHGTYSTV YSLGVGNTTA EKTYQENFLN KTGISTSGNSL DKEWVQDNLK SNLPGCLNG LQOTRCQTA CGENGGSRF GSSCDPDGCD FNSWGNKNTK NGIPNSDGM POKNDKNSG GKYGSCCSE MDIWEANSIC SAVTPHVCNDN LQOTRCQTA CGENGGSRF GSSCDPDGCD FNSWGNKNTK WGPGLIYDTH KPVQVTVTQF GSGGSYSTEIK RKYQOGKVI DSNNTNIAK SKOYNVSDE FCQQAQKAFG DNTSFTHKHGG QLGSALAKGM VLVLSLWLDHV YPTNSKKGAS DRGPQPTSG VPADVEKSAA DANVYSDIR FGAIIDSTY
SEQ ID NO : 127	MLAALALPTFA CSVGVGTKTTI EKTYQENFLN KTGISTSGNSL DKEWVQDNLK SNLPGCLNG LQOTRCQTA CGENGGSRF GSSCDPDGCD FNSWGNKNTK WGPGLIYDTH KPVQVTVTQF GSGGSYSTEIK RKYQOGKVI DSNNTNIAK SKOYNVSDE FCQQAQKAFG DNTSFTHKHGG DHDYDNMLWD SVPTNSKKGAS DRGPQCKTS SK
SEQ ID NO : 128	MEGFLLSFA LQFALEIGTO TSESHPSLTW ELNGARQSGQ IVDISNRWLV HDSGTTNCYD GNTWDSLCP DPEKCSQNCY LEGADYSGYT ATYGIQSGNT ALTPKFVTHG SYTNSNIGRL YLLKDESTYY VFOUNNKEFTFT FSTDVSCLPC GLNGALYFVE MDADGERSKY AGAKYGTGQV LGYCDAAQCPSS DLKPINGEANM SEGMWKPSGD KNAENGKYGGS CCSEMIDWES NSMATALPH VCKTIGQTRC SGKSECCGQD GODPAGNCND EDGCDEBNWWR MGDETEFRGK LVTDTKSFPV VVTQVTVTQF KPVQVTVTQF GSGGSYSTEIK RKYQOGKVI DSNNTNIAK SKOYNVSDE FCQQAQKAFG AFGTNDFKN KCGFPKLGKFS PSCRMNLVLS LMDHQVAML WLDSTPDTK DKSTPQFDVRG PCPTSSGRD DVEASASDATT VVYGMKFGA LDSTY
SEQ ID NO : 129	MLASLPLSN SLGTASMQAE THPLKLWTKN GGQTGYNGEVT VDSNWRWHTH KGSTKNCYDG NLWSKDLCPD AATCGKNCVLU EGADYSGYG YGITISGNOL KLGIVTYSY STNIGSYVYL LRDKNYEMF KLKNKEFTFT VDOSKLPCGL NGALYFYAMD EDGGSKNSI NKAGAQYGTG YCDACOPHDW KFINGEANVL DWKPQSDEN SGNGRGTACC TEMDWEANS MATAYPHVC TVTGLYRCEG TECGDNDQ RYNGICDKDG FSKALDTEV LVLSLWDHT ANMLWLDSTY PTDSALTAGS RGPCATSGSD PRDVEASAN ASVKFSDIKF GALDSTY
SEQ ID NO : 130	MLASLPLSN SLGTASMQAE THPLKLWTKN GGQTGYNGEVT VDSNWRWHTH KGSTKNCYDG NLWSKDLCPD AATCGKNCVLU EGADYSGYG YGITISGNOL KLGIVTYSY STNIGSYVYL LRDKNYEMF KLKNKEFTFT VDOSKLPCGL NGALYFYAMD EDGGSKNSI NKAGAQYGTG YCDACOPHDW KFINGEANVL DWKPQSDEN SGNGRGTACC TEMDWEANS MATAYPHVC TVTGLYRCEG TECGDNDQ RYNGICDKDG DENSYRMGDK SFPGQKTVD SSKPVTYTVTQF FITDNNODSG KLTETRKVV OGGRVINDSK VNLAGITAGN PTDPCDEA KKAFDNNDF EKKGGISALG TQLAGFLVLS LSLWDDHSVN MLWLDSYPT NASPAGLVE RGDAITSVF PADVEOSAD ASVTFDIKF GPIOSTY
SEQ ID NO : 131	MILALLVLGK SLGIATNQAE THPLKLWTRY QSKGSSTVN GEIVLDNSVR WTHHSGTINCY DNTWTSCLC DPPTCSNNC DLDDADYPTG KTGISTSGNSL KLGIVTYSY STNIGSYVYL LRDKNYEMF KLKNKEFTFT VDOSKLPCGL NGALYFYAMD EDGGSKNSI NKAGAQYGTG YCDACOPHDW KFINGEANVL DMKPOSDEN SGNGRGTACC TEMDWEANS MATAYPHVC TVTGLYRCEG TECGDNDQ RYNGICDKDG FNSWGNKNT FYGPGLIYDTH KSKFTVTVTQF VGNPVTETIK KYVONGKVIEN NSVTDTCNE QKKAFGDNND FEKKGGLGAL NMWLDDHSTP TNAAGAALGT ERGACATSSG KPSDVESQSP DATVTFSDIK FGPDIDSTY
SEQ ID NO : 132	MIGIVLQTV FGIGVGTQOQ EHPPLSLWQO CKGGGSTSV SGSTVLDNSW RWTHIPDGTG NYCDGDNWSS DLCPDPDTCSS NCYLEGADY SGTYGJ1STSG SSAKLGFVTK GSYTSTNIGR VYLLGDESHY KIFDLKNEF TFTVDDSENLE CGLNGALYFV AMDEGGASR FTIARKPGAKY

TABLE 7-continued

SEQ ID NO.	Amino acid sequence
SEQ ID NO : 133	MVFGTIVSFV YSIGVNTTA EHPKLTWKN GGSTINGEVT VDSNWRWHT KGSTKNCYDG NMSKDLCPD ATCGKNCVCL EGADYSGTYG DGCDDNSWRL GNOEFWGPOL IIDTGKEVTV VTOFVTKDGT DSGALSEIRR KYOGGKTE NTVKKGID EVDSITDEF C NOOQAFGDT NDFERKGGLS GLGRAFDGV VLVLSLWDHH DVNMMLWDSV YPTNPGKAG ADRGPCATSS GDPKEVEDY ASASVIFSDI KFGPDSTY
SEQ ID NO : 134	MLSFVFLGFV GVSLEIGTQQ SENHPTLWSQ QCTSSSSCQS QSGTIVLDSN WRNWHDSQT NYGDGNWESS DLCPDETC CS NCYLGDADY SGTYGITSNG SSKLKGFTVE GSYSTNLGSR VYLGKTFVHE TETVDVSNLP CGUNGALYEV EMEADGGKX YPLAKPGAQI GMGYCDAQCP HDMEPKINGA NVIDWKQET DEKTYOMFL NGKBEFTVTD VSOLPGCNG ALYFCVMDQ GGMSPRPDNQ AGAKYGTGTYC DAQCPDILKF INGLPNSDWT QKNDKNSNG MDIBRNWLS TAIVPTEGRV VQGTRCEGRV GCGNGDRF GSICDPDGCD FNSWRNGNKT FWGPGLLIDW KEPVTVYVTO IGSPTVTEIK RYVQGKTVL EYVQGKTVL NSYTMVNEGD KENSIKSIEKC TAQKTAEGDN DSFVKHGGEPS KLGGQSEPTKGQ VLVLSLWDHH TVNMLWLDV YPTNSKLGGS DRGP-CPTSSG VPADVEKNA DSSVVKSDIR FGSDISTYK
SEQ ID NO : 135	MLTLYVLLS LVSLEIGTQ QEBSHPAWTW QREGSSASGS IVLDSNMRWV HDSGTTNYCD GNEWSTDLCV SSTDTCQKCY IEGADYSGTY GITTSGSKLT LFVFTKGSKS TNIGSRYYL KDENTYETFK LKNKEFPTFY DDKSKLDCGIL GALYFVAMDA DGGKQKYSFF KPGAKYGMGY CDAQCPDHMK FISCKANTYD WFCQDNDENS GNGKLCTCS EMDIWEGNAK SQYTVHACT KSGQYEVETGQ DCGDSDSYRQ GTCKDGDGY ASYRMWGDHF YGECKTDTK QPTVTVQFI GPDPLTEIRL YIQQGKVINN SKTQNLSAVY DSITDAFCDA TKAASGDTND FKARGAMAGF SKNLDTPQVL VLSLWDDHTP TDSRDATAER GPCATSSGVP KDVESQNAQDA SVTFSDIKF AINSTSYN
SEQ ID NO : 136	MFGFLLSLFA LQFALEIGTQ TSESHPSITW ELNGARQSQ IVDISNMRWV HDSGTTNYCD GNTWSIDLCP DPEKCSONCY LEGADYSGTY GISACGSQLT LQFPTVKGSKS TNIGSRYYL KDENTYETFK LKNKEFPTFY DVSNLPGCIN GALYFVAMDA DGGKAKYPLA KPGAKYGMGY CDAQCPDHMK FINEBANVYD WFCQDNDENS GNGKLCTCS EMDIWEGNAK SQYTVHACT KNSARCBETEC GDDSAQSYRN GICIKDGDGF NSWRNGNKTW FGPLTVDSS KPVTVQFI GPDPLTEIRI WQGGKVION SFTNVSIGTS VDSITNTFCD ESKVATGDTN DFKAKGMSG FSKALDTEVV LVLSLWDDHT ANMLWLDSTY PSNSTAIGAT RGPATSSG PKNVEASAN ASVFKSDIK GADSTY
SEQ ID NO : 137	MLALVYFLLS LVSLEIGTQ OSEDHPKLWV QNGSSVSGS IVLDSNMRWV HDSGTTNYCD GNLWSTDLCP SSTDTCSKCY IEGADYSGTY GITSGSKVY LFVFTKGSKS TNIGSRYYL KDENTYETFK LKNKEFPTFY DDSQLNQGILN GALYFVAMDA DGGKAKYSSF KPGAKYGMGY CDAQCPDHMK FISCKANTYD WFCQDNDENS GNGKLCTCS EMDIWEGNAK SQYTVHACT KSGQYEVETGQ QCGDSDGDR FKGTCDKDGC DYASPRWGDQ SFYGBGKTVD TQOPTVTVQ FIGDPLTEIR RLYVQGKTVI NNSKTSNLAD TYDSITDKFC DATEKZGDT NDFPAKGAMS GFSTNLNTAQ VLVLSLWDHH TANMLWLDST YPTDSTKGA SRGP-CAVLSG VPDKVNEVSQG SAQVYSDIK FGAINSTSY N
SEQ ID NO : 138	MLALVYFLLS FVSVLEIGTQ OSEDHPKLWV QNGSSVSGS IVLDSNMRWV HDSGTTNYCD GNLWSTDLCG SSTDTCSKCY IEGADYSGTY GISACGSQLT LQFPTVKGSKS TNIGSRYYL KDENTYETFK LKGKEFPTFY DVSNLPGCIN GALYFVAMDA DGGKAKYSSF KPGAKYGMGY CDAQCPDHMK FISCKANTYD WFCQDNDENS GNGKLCTCS EMDIWEGNAK SQYTVHACT KSGQYEVETGQ QCGDSDGDR FKGTCDKDGC DYASPRWGDQ SFYGBGKTVD TQOPTVTVQ FIGDPLTEIR RLYVQGKTVI NNSKTSNLAD YDTSITDKFC DATEKZGDT NDFPAKGAMS GFSTNLNTAQ VLVLSLWDHH
SEQ ID NO : 139	MFLALFVLGK SLGATNQAE NHPKLTWTR QSGKSGCVN GEVYLDVTVN WTHHSGTVN DGTNTWTSCL PDPCGSSNC DLDGADYPST YGISSGNSL KLGFTVTHGSY STNIGSRYYL LRDSKNEYMP KLRKNEFPTF VIDSKLPCGL NGALYFVAME EDGGVAKNSI NKAQAYGTG YCDAQCPDHM KPTNGEANVY DMKPOSDNEN SGNGRGAAC IEMDIWEANANS MATAVPHVC TVTGIIHCEG TECGDTDANO RYNG1CDKG CDENSYRMGD KSFRGVGKTVV DSSKPVTVTQ QFTVTSQGDD GTLSEIIRKY VQGKVIEINS KVNIAGITAV NISITDFCNE QKAFGDDND FEKKKGSGGAL SKQDLMGHML VLSLWDDHV NMWLWLDSTY TDAAAGAL GT ERGACATSSG KPSDVEQSP DASVYSDIK FGAPIDSTY
SEQ ID NO : 140	MLCLLISIAN SLEGYGTNTAE NHPKLTWTR QSGKSGCVN LMSDKYCOPDA ATCGKNCVIE GADYQGTYGV SSSGPDGLTLT FVTHGQYSTN VGSRLYIMKD EKTYQVNMLN GKEFVTVN GKEFVTVN GMAKVDNQA GAKYGTGCD AQCPDILKF INGLPNSDWMK PQKNDKNSGN GKYGSCSSEM DIWFAASQAT AYIPHCDKL EQTRCSGSSC GHTGGERFS SSUPDGCDF NSWRNGNKTW WGPLVLDHK RPVQVTVQFV GSGNISTEIK RKYVQGKTVI DMSMSNIAIM SKQYNSVSDD FCQAQKAFG DNDSPTKHGG DHDVNLWLD SVTFPSNKP GSDRGPCCKTS SGIPADVESQ AASSSVKYS IREGAIDSTY K

TABLE 7-continued

SEQ ID NO.	Amino acid sequence
SEQ ID NO : 141	MLCIGLISFY YSLGVGNTAA ETPHKLTWKN GGTGVNGEVT VDSNWRWHTH KGSTKNCYDG NLWSKDLCPD AATCGKNCVLC GLADYSGTYG VTSQSLRLMK KFVHGQSTV TNGSRLYLMK DEKTIQFLNL NGKEFTVTD VSNLPCTG ALYHVNED GGTRKPDNE AGAKTGTYC DAOCPDILK INGPNSDGM KPOSNDKNSG NGKYGCSCSE MDIWEANSIC SAVTPHVDT LQOTRQGTA CGENGGSRF GSSCDPDGCD FNSWRGNGKT FYGPGLIVDT KSKPTVTVTQF VGSPTVTEIKR KYVQNGKVI NESPNSIEGD M KENSIISDFC TAQKAFGDT DSFTKHGGFK QLGSLAKGM VLVLSLWDDH YPTNSKAKGS DRGPCEPTSG VPADVESETSA NANVYSDIR FGAIUTSYK
SEQ ID NO : 142	MUCCLGIAA SL DAGNTAA NHPLSLWKG GSSVSGSVTV DANEFWTHIK GETKNCVGN LMSDKYCPDA ATCGKNCVIE GADYQGTYEV SASGNALTLT FVTHGQSTVN VGSRLYLLKD EKTYQIFNLI GKETFTVTD SNLPCGNGA LYFVQMDAGD GTAK5DNKA GAKYGTGTYD AQCPTDILK INGLPNSDGMK POKNDKNSG GRGSOCSEM DVWEANSLAT AYTPHVDCKL EQYRCDRAC GONGGDRFS SSCPDGCDP FNSWRGNGKTW WGPFLIVDTK RQVYQGGKVI QPVQVTVQWV SGYPTNSKPK VLSLWDDH DNEPTKLSDH TQYQNSVYSD FCVAQRKAFG DNDSTKHGG FRQLGATLAK GHVLVSLWDDH DHDVNMLWDDH SGYPADEVSQ GADRGPCKTS K
SEQ ID NO : 143	MLGIGRVCTV YSLGVGNTAA ENHPKLTWKN SGSTTINGEVT VDSNWRWHTH KGSTKNCYDG NLWSKDLCPD AATCGKNCVLC GLADYSGTYG VTSQSLRLMK KFVHGQSTV TNGSRLYLMK DEKTIQFLNL NGKEFTVTD VSNLPCTG ALYHVNED GGTRKPDNE AGAKTGTYC DAOCPDILK INGPNSDGM KPOSNDKNSG NGKYGCSCSE MDIWEANSIC SAVTPHVDT LQOTRQGTA CGENGGSRF GSSCDPDGCD FNSWRGNGKT FWGPGLIVDT KEPVTVTVTQF VGSPTVTEIKR KYVQNGKVI NESPNSIEGD M KENSIISDFC TAQKAFGDT DSFTKHGGFR QLGOSRFTKGQ VLVLSLWDDH YPTNSKAKGS DRGPCEPTSG VPADVESETSA GSSVSKYSDIR FGSIUTSYK
SEQ ID NO : 144	MATLVGILV LEALEVALII GTQTSSESHPS LSWELNGQRO TGSPVIVDSNW RWLHDSGTTN CYDGNIENSSD LCPDPPEKCSQ NCYLEGADYS GTYGISSSGN SLOQGFVTKG SYTSNIGSRV YLLKDNTYA TFKLKRNKEFT FTADVSILPC GLNGALYFVA MPADGEKSKY PLAPFGAKTG MGYCDAQCPH DMKPINGEAN ILDWKPSND ENAGAGRGT CCTEMDWEA NSQATAYTVH ACSKNARCEG TECCDDGDRY NGICDKDGD FNSWRGNGKT FERGNLIVDS IYVQYGGKTI RQVYQGGKTI NSEFTNSIAD FC NENKVATGDT NDFTAKGMS GFSKALDTEV VLVLSLWDDH TANMMLDST YPTDSALCA SGPCTATSG EPKDVESASA NASTPKSDIK FGADSYT
SEQ ID NO : 145	MLTVYFLL LVVSLLEITQ QSESHPSLQW QNGSSSVGS IVLDNSWRWV HDSGTTNCYD GNLSWSTDLC P SSDTCTSKCY LEGADYSGTY GITSGSGLT LKFVTKGYS TNIGSRYL KDENTYETFK LKNEFETFTV DDESKLDCGNN DGGKAKYSSP KPGAKYGMGX CDAQCPHDMK FISGKANDD WKPQNDENS GNGKLGTS EMDIWEANSIC SQAFTVHACT KSGQVTCQO OCQDTSGDR FKGTCDKDGC DYASVPRWDMK SFYCEGKTVD TKEPLTVTU FVGDPLTEIR RVYQYGGKTI NNSEFTNSIAD DATKESGDT NDFTAKGANS GSTLNTAQ VLVLSLWDDH TANMMLDST YPTDSITKGA SRGP CAVSSG VPDVSEQHSQ DATVYSDIK PGAINSTFKW N
SEQ ID NO : 146	MLSLSVIFL GLGBSLGVT QOSESHPSL WQNCEAKGSC QSVSCTVLD SMTWRLHSD TINCYDNEW STDLCEDAST CDKNCYIEGA DYSCTGIGTS SGAOLKLGFV TKGSYSTNG SRVYLIRDHS HYOLFKLNH EFTFTVDSQ LPCGLNGLY FVEMADGGA KPGQYGMGX CDAQCPHDMK FITEBANVKD WKPQETDENA SNGHYGACT EMDIWEANSIC ATAYPHICS KTGYRCBTG EGCDNDANQR YNGCDDKGCF DFNSTRLGNK TFPGLTVD SNKAMIVTU FTTSNNODSG ELSEFTVLDQ OGQTQNSID TIVQGTTN KITOFCDET KVTFRDNDP KAKGEGSGLS KSLESAGLV LSLSWDHSVN MLWLDSTPT DSAGKPGADR GPAITSGDP KDVESQSPNA SVTFSDIKFG PIDSTY
SEQ ID NO : 147	MILALLVGLK SLGIATNQAE TRPKLWTRY QSKGSSSTVN GEIVLDSNWV WTHHSNTCY DENTWTSCLC DPPTCSNNC DLDGADYPT YGISTSGNSL KLGPFVTHSY STNIGSRYL LKDTRSYEM KLNKNEFETFT VDSSKLPCL NGALYVAMD EDGTSKNSI NKAGAQYGTG YCDAQCPHDMK KFINGEANVW DNKPQSVTNDEN SGNQGTYGAC MATAVPHCS TVTQFTRREG TECGTDNDQ RYNGICDDRG ODENEYRLGD KSFPGVQKTV DSSKPVTVTQF QFTVTSNGDS GTLSFIRKRY VOGSKVTHNS KVNGVATGAT NSVTDTCNE OKRAFDNDN PEEKGGFGAL SKQLVAGMVL VLSLWDDHSV NMWLWLDSTP TNAAGAQLGT ERGACATSGS KPSDVESQSP DATVTFSDIK FGADSYT
SEQ ID NO : 148	MLCVGLFGLV YSLGVGNTAA ETPHKLTWKN GGTGVNGEVT VDSNWRWHTH KGSTKNCYDG NLWSKDLCPD AATCGKNCVLC GLADYSGTYG SGTYGUTSSG DSUTLKETHY GSYSTNGSR LYLLKDNTY QIENLAKEE TETVDSMLP CUNGLALYFV EMQDGGKFK HKEAEAGKY GTGYCDAQCP TDLKFDGIA NSDGWKRQDN DENSENGKYG SCSCENDIVE ANSIAVATP HVCDTQKQR CGTIAQENG GDFTGSECD PDGCDENNSW QGNKSEWFGP LIIDTKTSVQ VVTOFQTSGS SVTETRKVY QNCKVNLSY STLSGTEKYN SISDVCNAQ KKAFGDTNSF ENHGGFKRFV QHJODMVLYL SWDDDHVNM LMLDSVTPN SNKGADRG P CETSGVPTD VESKSASV KYSDRFGPI DSTYK
SEQ ID NO : 149	MLLCUMSIAY SLGVGNTAA NHPLSLWKG GSSVSGSVTV DANEFWTHIK GETKNCVGN LMSDKYCPDA ATCGKNCVIE GADYQGTYEV SASGDBGLLT FVTHGQSTVN VGSRLYLMK EKTYQIFNLN GKEFTFTVTD SNLPCGNGA LYFVQMDAGD GMAYKPDNOA GAKYGTGTYD

TABLE 7-continued

SEQ ID NO.	Amino acid sequence
	AGCPTDLKF INGLPNSDGWIK POKNDKNSGN GKYGCCSEM DIWEANSQAT AYTPHVCDDL EQRTRCSGSA CHTGGGERFS SSCCDPDGCDF NSWRIGCNKTF WGPLIIVDTIK RPKVQVNTK RVKYGKGVII DNEMSNAGM TKOYNSTSD FCOAQKAKF DNDSEFTKHGG FRQLCATLGK GHVLVLSLND SVIPTINSNPK GSDRGPCRTS SGIPADVESQ ASSSVTKSD IRFGAIDSTY K
SEQ ID NO : 299	QSACTLQSET HPPLTWOQCS SGGTCTQOTG SVVIDANWRW THATNSSTNC YDENTWSSTL CPDNETAKN CCLDGAAYAS TYGYTTSGNS LSIGRVITOSA QKNGARLYL NASDTTYQEF TLLGNEBESFD VDVSLQPCGL NGALYFVND ADGGVSKPT NTAGAKYGTG YCDSOCPRDL KFINGQANVE GWEPSNNAN TGIGGHGSC SEMDIWEANS ISEALTHPHC TTGQELICRS TTVQETKSDN AYGGTCPDG CDNPYRILGN TSFYCPGSSP TLJDTTKLKV VTOFNETSAAI NRYVYONGVT FQOPNAELGS YSGNELNDYY CTAEAEAFGG SSESFKKGGLT QFKATSGEM VLVMSLWDDY YANMILWLDST YPTNETSSTP GAVRGSCSTS SGVPAQVEQSQ SPAKVTFSN IKFGPIGSTG NPGGNPPGG NPPGTTTRR PATTGSSPG PTQSHYHGCGC AGTGTQVNL PYYSQCL
SEQ ID NO : 300	QSACTLQSET HPPLTWOQCS SGGTCTQOTG SVVIDANWRW THATNSSTNC YDENTWSSTL CPDNETAKN CCLDGAAYAS TYGYTTSGNS LSIGRVITOSA QKNGARLYL NASDTTYQEF TLLGNEBESFD VDVSLQPCGL NGALYFVND ADGGVSKPT NTAGAKYGTG YCDSOCPRDL KFINGQANVE GWEPSNNAN TGIGGHGSC SEMDIWEANS ISEALTHPHC TTGQELICRS DCGCQTSDN RYGGTCPDG CDNPYRILGN TSFYCPGSSP TLJDTTKLKV VTOFNETSAAI NRYVYONGVT FQOPNAELGS YSGNELNDYY CTAEAEAFGG SSFSDRKGGLT QFKATSGEM VLVMSLWDDY YANMILWLDST YPTNETSSTP GAVAGSCSTS SGVPAQVEQSQ SPAKVTFSN IKFGPIGSTG NPGGNPPGG NPPGTTTRR PATTGSSPG PTQSHYHGCGC AGTGTQVNL PYYSQCL
SEQ ID NO : 301	MSALINFNYY KSALILGSLV ATAGAQOQIT YTAEHTPLS WSTCKSGSC TTNSGAIITD ANMRWVGVN TSTMCYTGNT WNTAICDTDA SCAQDCALDG ADYFSGTGTL TSGNSLRLNP VTGSNNGSRT YLMADNTHYQ TEDLNLQEQT FTVDVSHIPC GLNGALYEVY MDAGGVSKY PNNKZAGAQYVG VGYCDSOQPCR DLKPIAGAN VEGWTSSNN ANTIGLNGHA CCABLDIWA NIISEALTHPH PCDTPLSVC TTDAAGGVTYS SDKYAGTCDP DGCDENPVRLL GTDTPYGSK TVDTTPKITY VTOFVTDGTT STGTSLSEIRR YYVQNGVTP QPSKSLSGVS GNVINSDFCD AEIISFGETA SFSEHGGIAK MEGAMEAGMV LVMSLWDDY VNMLWLDSTY PTNATGPRA AKGSCPTTSG DPKTVBQSQG SSYTFSDIR VGPBNSTPSG GSSTGGSSTT TASGTTKA SSISSTSSTP GTCHVAHWQ CGCGQGTTGT TCASOTTCV VNFYSSQCL
SEQ ID NO : 302	QIIGTYTAET HPSLSWSTCK SGGSCTTNS AITLDANWRW VHGYNTSTNC YTGTNTATAI CDTDASQAQ CALDADYSG TYGYTTSGNS IRLNFTVGEN VGSERTYLWD NTHYQIFDLV NOEFTTVDV SHLPGLNKA LYFTVMDADG CVSKYPIANKA GAQYGYGYCD SQCRDLKEFI AGQANVGEWI PSSENNTANGL GHGACCAEL DIWEANSIE ALTHPDCDP GLSYVCTDAC GTYISSDKYA GTCDDPGCDF NPYRLGVTFD YSGGKTVDTT KPTIVVTOFV TDGTSTGTL SEIRRYYQN GVVIQPOSSK ISGSGNAYIN SDFCDALST FGETAFSKH GGLARMAGM EAGMVLYMSL WDDYSVNMLW LDSTYPTNAT GTPGAAKGS C PTTSGDPKTY ESQSGSSTVT FSDIRVGPEN STSGGSSTG GSSTTTASGT TTTKASSTST SSSTSTGTVA AHWGQCCQG WTGPTCAASG TTCTVVNPYY SQCL

## SEQUENCE LISTING

The patent contains a lengthy "Sequence Listing" section. A copy of the "Sequence Listing" is available in electronic form from the USPTO web site (<http://seqdata.uspto.gov/?pageRequest=docDetail&DocID=US09096871B2>). An electronic copy of the "Sequence Listing" will also be available from the USPTO upon request and payment of the fee set forth in 37 CFR 1.19(b)(3).

What is claimed is:

1. A polypeptide comprising a variant cellobiohydrolase I (CBH I) catalytic domain as compared to a reference CBH I catalytic domain, wherein the variant CBH I catalytic domain comprises an amino acid sequence having at least 90% sequence identity to amino acids 26-455 of SEQ ID NO:1, the polypeptide comprising:
  - (a) a substitution at the amino acid position corresponding to R268 of *T. reesei* CBH I of SEQ ID NO:2 (R268 substitution);
  - (b) a substitution at the amino acid position corresponding to R411 of *T. reesei* CBH I of SEQ ID NO: 2 (R411 substitution); or
  - (c) both an R268 substitution and an R411 substitution, wherein substitution (a), (b) or (c) decreases product inhibition as compared to the reference CBH I catalytic domain.
2. The polypeptide of claim 1, which comprises an R268 substitution.
3. The polypeptide of claim 2, wherein the R268 substituent is a lysine.
4. The polypeptide of claim 2, wherein the R268 substituent is an alanine.
5. The polypeptide of claim 1, which comprises an R411 substitution.
6. The polypeptide of claim 5, wherein the R411 substituent is a lysine.
7. The polypeptide of claim 5, wherein the R411 substituent is an alanine.
8. The polypeptide of claim 1, wherein the variant CBH I catalytic domain comprises an amino acid sequence having at least 95% sequence identity to amino acids 26-455 of SEQ ID NO:1.
9. The polypeptide of claim 8, wherein, other than said R268 and/or R411 substitutions, the variant CBH I catalytic domain comprises the sequence of amino acids 26-455 of SEQ ID NO:1.
10. The polypeptide of claim 1, wherein the variant CBH I catalytic domain comprises one of the following amino acid substitutions or pairs of amino acid substitutions as compared to a protein of SEQ ID NO:1:
  - (a) R273K and R422K;
  - (b) R273K and R422A;
  - (c) R273A and R422K;
  - (d) R273A and R422A;
  - (e) R273A;
  - (f) R273K;
  - (g) R422A; and
  - (h) R422K.
11. The polypeptide of claim 1, wherein the variant CBH I catalytic domain comprises the amino acid substitutions R273K and R422K as compared to a protein of SEQ ID NO:1.
12. The polypeptide of claim 1, wherein the variant CBH I catalytic domain does not comprise both R273K and R422K substitutions as compared to a protein of SEQ ID NO:1.
13. The polypeptide of claim 1 in which the variant CBH I catalytic domain is linked to a cellulose binding domain.
14. The polypeptide of claim 13 in which the catalytic domain is linked to a cellulose binding domain via a linker.
15. The polypeptide of claim 13 in which the cellulose binding domain is C-terminal to the catalytic domain.
16. The polypeptide of claim 13 in which the cellulose binding domain is N-terminal to the catalytic domain.
17. The polypeptide of claim 1 which further comprises a signal sequence.
18. The polypeptide of claim 1 towards which cellobiose has an IC<sub>50</sub> that is at least 2-fold the IC<sub>50</sub> of a reference CBH I lacking said R268 substitution and/or R411 substitution.
19. The polypeptide of claim 1 which CBH I activity that is at least 50% the CBH I activity of a reference CBH I lacking said R268 substitution and/or R411 substitution.
20. A composition comprising a polypeptide according to claim 1.
21. The composition of claim 20 in which said polypeptide represents at least 1% of all polypeptides in said composition.
22. The composition of claim 21 in which said polypeptide represents at least 5% of all polypeptide in said composition.
23. The composition of claim 22 in which said polypeptide represents at least 25% of all polypeptide in said composition.
24. The composition of claim 20 which is a whole cellulase.
25. The composition of claim 24, wherein the whole cellulase is produced by a host cell that recombinantly expresses said polypeptide.
26. The composition of claim 20 which is filamentous fungal whole cellulase.
27. A fermentation broth comprising a polypeptide according to claim 1.
28. The fermentation broth of claim 27, which is a filamentous fungal fermentation broth.
29. The fermentation broth of claim 27, which is a cell-free fermentation broth.
30. A method for saccharifying biomass, comprising: treating biomass with a composition according to claim 20.
31. The method of claim 30, further comprising recovering monosaccharides.
32. A nucleic acid comprising a nucleotide sequence encoding the polypeptide of claim 1.
33. A vector comprising the nucleic acid of claim 32.
34. The vector of claim 33 which further comprises an origin of replication.
35. The vector of claim 33 which further comprises a promoter sequence operably linked to said nucleotide sequence.

**161**

- 36.** The vector of claim **35**, wherein the promoter sequence is operable in yeast.
- 37.** The vector of claim **35**, wherein the promoter sequence is operable in filamentous fungi.
- 38.** A recombinant cell engineered to express the nucleic acid of claim **32**.
- 39.** The recombinant cell of claim **38** which is a eukaryotic cell.
- 40.** The recombinant cell of claim **39** which is a filamentous fungal cell.
- 41.** The recombinant cell of claim **40**, wherein the filamentous fungal cell is of the genus *Aspergillus*, *Penicillium*, *Rhizopus*, *Chrysosporium*, *Myceliophthora*, *Trichoderma*, *Humicola*, *Acremonium* or *Fusarium*.
- 42.** The recombinant cell of claim **40**, wherein the filamentous fungal cell is of the species *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Penicillium chrysogenum*, *Myceliophthora thermophila*, or *Rhizopus oryzae*.
- 43.** The recombinant cell of claim **39** which is a yeast cell.
- 44.** The recombinant cell of claim **43** which is a yeast cell of the genus *Saccharomyces*, *Kluyveromyces*, *Candida*, *Pichia*, *Schizosaccharomyces*, *Hansenula*, *Klockera*, *Schwanniomyces* or *Yarrowia*.
- 45.** The recombinant cell of claim **44**, wherein the yeast cell is of the species *S. cerevisiae*, *S. bulderi*, *S. barnetti*, *S. exiguis*, *S. uvarum*, *S. diastaticus*, *K. lactis*, *K. marxianus* or *K. fragilis*.
- 46.** The recombinant cell of claim **45**, which is a *S. cerevisiae* cell.
- 47.** A host cell transformed with the vector of claim **33**.
- 48.** The host cell of claim **47** which is a prokaryotic cell.
- 49.** The host cell of claim **48** which is a bacterial cell.
- 50.** The host cell of claim **47** which is a eukaryotic cell.
- 51.** A method of producing a polypeptide according to claim **1**, comprising culturing the recombinant cell engineered to express said polypeptide under conditions in which the polypeptide is expressed.

**162**

- 52.** The method of claim **51**, wherein the polypeptide comprises a signal sequence and wherein the recombinant cell is cultured under conditions in which the polypeptide is secreted from the recombinant cell.
- 53.** The method of claim **52**, further comprising recovering the polypeptide from the cell culture.
- 54.** The method of claim **53**, wherein recovering the polypeptide comprises a step of centrifuging away cells and/or cellular debris.
- 55.** The method of claim **53**, wherein recovering the polypeptide comprises a step of filtering away cells and/or cellular debris.
- 56.** A method for generating a product tolerant variant CBH I polypeptide, wherein the variant CBH I catalytic domain comprises an amino acid sequence having at least 90% sequence identity to amino acids 26-455 of SEQ ID NO:1 comprising:
- (a) modifying the nucleotide sequence of a CBH I-encoding nucleic acid so that the nucleic acid encodes a variant CBH I polypeptide, wherein said variant CBH I polypeptide comprises:
    - (i) an R273 substitution of SEQ ID NO: 1;
    - (ii) an R422 substitution of SEQ ID NO: 1; or
    - (iii) both an R273 substitution and an R422 substitution; and
  - (b) expressing said variant CBH I polypeptide, thereby generating a product tolerant variant CBH I polypeptide.
- 57.** A method for producing ethanol, comprising:
- (a) treating biomass with a fermentation broth according to claim **27**, thereby producing monosaccharides; and
  - (b) culturing a fermenting microorganism in the presence of the monosaccharides produced in step (a) under fermentation conditions, thereby producing ethanol.

\* \* \* \* \*